

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 2004, 00:43:16 ; Search time 4101.31 Seconds
(without alignments)
10987.223 Million cell updates/sec

Title: US-10-009-254-1
Perfect score: 1509
Sequence: 1 atgaaaagaagaatgattca.....gtcgtcggttaacgtttcttaa 1509

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hcc:*
9: gb_estl:*
10: gb_est2:*
11: gb_hcc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rpd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	76	5.0	843	28	BZ643413
2	69.8	4.6	844	29	BX139987
3	69.4	4.6	494	29	FR0048073
c 4	69	4.6	829	29	BX173672

5	68	4.5	989	29	CNS02HA4	AL197365 Tetraodon
6	67.8	4.5	641	28	AQ946120	AQ946120 Sheared D
c 7	67	4.4	619	29	CG372306	CG372306 OGA4B108TC
8	66.4	4.4	501	29	FR0048173	AL444958 Fugu rubr
9	66.4	4.4	599	13	BX729701	BX729701 BX729701
10	66.2	4.4	575	28	BZ780846	BZ780846 i120h11.9
c 11	66.2	4.4	1101	29	CNS012TP	AL102007 Drosophil
c 12	65.2	4.3	1101	29	CNS0039G	AL063921 Drosophil
c 13	64.6	4.3	735	29	CNS04NSM	AL299119 Tetraodon
c 14	62.8	4.2	480	28	BZ643398	BZ643398 OGA0W57TC
15	62.6	4.1	513	28	BZ423698	BZ423698 iD52c10.9
c 16	61.8	4.1	994	13	BX414650	BX414650 BX414650
c 17	61.4	4.1	470	29	FR0018463	AL011359 F.rubripe
c 18	61	4.0	450	29	FR0025683	AL018519 F.rubripe
c 19	60.8	4.0	479	29	BH211305	BH211305 Sm1-30G22
20	60.6	4.0	788	28	AZ183942	AZ183942 SP_1002.A
21	60	4.0	525	28	BZ351620	BZ351620 hW04F06.G
c 22	60	4.0	806	29	CNS04ABE	AL281759 Tetraodon
c 23	59.8	4.0	773	29	CNS01VTG	AL169549 Tetraodon
24	59.6	3.9	617	12	BJ391130	BJ391130 BJ391130
c 25	58.6	3.9	427	28	BZ423231	BZ423231 iD52c10.b
26	58.6	3.9	707	29	CG403581	CG403581 ZMWBB024
c 27	58.4	3.9	914	28	AZ205202	AZ205202 SP_0100.A
c 28	58	3.8	548	28	BH200391	BH200391 Sm1-39D16
c 29	58	3.8	619	29	FR0006944	Z90754 F.rubripes
30	57.6	3.8	641	29	CE682324	CE682324 tigr-988-
31	57.4	3.8	619	29	FR0047601	AL444386 Fugu rubr
c 32	57.4	3.8	781	29	BX145762	BX145762 Danio rer
c 33	56.8	3.8	587	28	BH873429	BH873429 hp45a07.b
c 34	56.8	3.8	856	28	BH178274	BH178274 011_F_07-
c 35	56.8	3.8	856	28	CNS07KWL	AL615231 T7 end of
c 36	56.8	3.8	1200	13	BX437758	BX437758 BX437758
c 37	56.6	3.8	713	12	BJ408597	BJ408597 BJ408597
c 38	56	3.7	541	12	BJ406947	BJ406947 BJ406947
39	56	3.7	854	28	AQ411842	AQ411842 CP08065B
c 40	55.8	3.7	363	29	CG807311	CG807311 1118080F0
c 41	55.8	3.7	670	28	BZ388480	BZ388480 EINDL21TF
c 42	55.6	3.7	471	29	CG807194	CG807194 1118078H0
c 43	55.4	3.7	933	28	AZ204694	AZ204694 SP_0100.A
c 44	55.2	3.7	350	28	BH879665	BH879665 ht47h12.9
c 45	55.2	3.7	655	12	BJ403823	BJ403823 BJ403823

ALIGNMENTS

RESULT 1
BZ643413/c
LOCUS BZ643413 843 bp DNA linear GSS 29-JAN-2003
DEFINITION OGA0W57TM ZM_0.7_1.5_KB Zea mays genomic clone ZMWBA011J17,
genomic survey sequence.

ACCESSION BZ643413
VERSION BZ643413.1 GI:28104915
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays

REFERENCE 1 (bases 1 to 843)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T., Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
TITLE Consortium for Maize Genomics
JOURNAL Unpublished (2002)
COMMENT Other_GSSs: OGA0W57TC
Contact: Cathy Whitelaw
TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR

Class: sheared ends.
Location/Qualifiers
1.843
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMB0111J17"
/note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

FEATURES
source
ORIGIN
Query Match 5.0%; Score 76; DB 28; Length 843;
Best Local Similarity 43.7%; Pred. No. 6.6e-07;
Matches 334; Conservative 0; Mismatches 430; Indels 0; Gaps 0;
Qy 171 TAATGCAAAATGATCTGATTCGAATGAAGATGGAGCTTCATTATTAATTCCTCAAGGTAA 230
Db 778 TACTGCTGCTGATGCTGCTACTGCTGCTGCTACTGCTGCTACTGCTACTGCTACTGCTGCTG 719
Qy 231 AGAAGCTGAGTATAAAGCTTCAACTGATTTTAAATTCCTTTTACGCAACTACTATG 290
Db 718 TACTGCTACTACTGCTGCTACTACTACTGCTACTACTGCTACTACTGCTACTACTGCTACTG 659
Qy 291 AGGAGAACATATGTAACATAAAGATGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 350
Db 658 TGCTACTACTACTACTGCTACTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 599
Qy 351 TAAATCTATATACCTAATATACACAGTTTCACATGTTACTGAGTCAATATGATG 410
Db 598 TGCTACTCCAGCTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 539
Qy 411 TACTGAGTTTAAATGTTTCCCAATATGGATATTAATTTCTTTTCTGAGCTGTTAATAA 470
Db 538 TGCTGCTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 479
Qy 471 TGGAGCTGTAAATATGTTTACATCTGAATCCAAATGCTACTATTCATGAAAAGATAC 530
Db 478 TGCTGCTGCACTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 419
Qy 531 TGATGCACTGCGGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 590
Db 418 TACTGCTAATGCTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 359
Qy 591 TGATACAGTCAAAATATATCTATTAAGATGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 650
Db 358 TACTACTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 299
Qy 651 AGGTATCAATATGTTTAAAGGATATATGCCATCTGCTTCTGTAGTTGATTTGAACGA 710
Db 298 AGCTACTACAGTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 239
Qy 711 AGGTCTTATGAAGTACTATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 770
Db 238 TACTGCTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 179
Qy 771 TTCGGAAGAAACCACTGGGAGTATAACCTGTTAGAGGAAAATAATATTTTCAGATTAC 830
Db 178 TGCTACTGCTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 119
Qy 831 TATTTCGCTGGGAGCTACCAATACTCCAAACGGGAATACTCAAAATGAGGACTAATGATGA 890
Db 118 TAATACTGCTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 59
Qy 891 CTTTCTTTTATAAGGGAATAAATAACAATCAAGTCACTTATACAG 934
Db 58 TACTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 15

FEATURES
source
ORIGIN
Query Match 4.6%; Score 69.8; DB 29; Length 844;
Best Local Similarity 46.3%; Pred. No. 1.8e-05;
Matches 230; Conservative 0; Mismatches 267; Indels 0; Gaps 0;
Qy 137 ATAAAGCATATAAGTTTTTGATGCAGAAATAGATATGCAATGATCTGATTGCGAATA 196
Db 348 ATAAATTAATAATGATAATAATAATGATAATAATAATGATAATAATAATGATAATAATA 407
Qy 197 AAGATGGAGCTCTCTTATTTAAATTCCTCAAGGTAAAGAGCTGAGTATAAAGCTTCAACTG 256
Db 408 TTGATAATAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 467
Qy 257 ATTTTAAATCTCTTTTACGCAACTACTATGAGGAGGAGACATATGTAACCTAAAAAAG 316
Db 468 ATAATAATAATAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATG 527
Qy 317 ATACTGCTGAGCAAAATGAGATTGCGACATGGGCTAAATCTATATCAGCTAACTACTACAC 376
Db 528 ATAATAATAATAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 587
Qy 377 CAGTTTCCACTGTTACTGAGTCAAAATAATGATGCTGAGGTTATTAATGTTTCCCAAT 436
Db 588 ATGATGATAATAATAATAATGATAATAATAATAATAATAATAATAATAATAATAATAATA 647
Qy 437 ATGGATATTAATGATGTTTCTAGCACTGTTAAATGAGGCTGTAATATGTTGTTACATCTG 496
Db 648 ATAATAATAATAATGATGATAATAATAATAATAATAATAATAATAATAATAATAATAATA 707
Qy 497 TAACTCCAAATGCTACTATTTCATGAAAAGAAATACTGATGCGACATGGGAGATGGTGGT 556
Db 708 CTAATAATGCAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 767
Qy 557 GAAAACTGTGATGCAAAAAACGCTACTCGGTTGGTGATACAGTCAAAATATATCTATTACT 616
Db 768 ATAATAATGATACTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 827
Qy 617 ATAAAGATGCACTCAAT 633
Db 828 ATAATAATGATGATAAT 844

RESULT 3
BX139987
LOCUS

RESULT 2
BX139987
LOCUS

FR0048073 494 bp DNA linear GSS 05-JAN-2001

FR0048073 844 bp DNA linear GSS 13-MAR-2003

DEFINITION Fugu rubripes GSS sequence, clone 263K15b08, genomic survey sequence.
ACCESSION AL444858
VERSION AL444858.1 GI:12052694
KEYWORDS GSS; genome survey sequence.
SOURCE Takifugu rubripes (Fugu rubripes)
ORGANISM Takifugu rubripes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodontidae; Takifugu.
REFERENCE 1 (bases 1 to 494)
AUTHORS Clark, M.S.
TITLE Direct Submission
JOURNAL Submitted (29-SEP-2000) MRC Human Genome Mapping Project Resource Centre Hinxton, Cambridge, CB10 1SB, UK Email: biohelp@hgm.mrc.ac.uk
COMMENT Vector: pBluescript II KS
PRIMER: KS
DESCR: One pass dye-terminator sequencing of BAC (pBelobACII) cloned genomic sequence
The BACs can be obtained from <http://www.incyte.com>.
FEATURES
source 1..494
Location/Qualifiers
/organism="Takifugu rubripes"
/mol_type="genomic DNA"
/db_xref="taxon:31033"
/clone="263K15b08"
/clone_lib="BAC 263K15"
ORIGIN
Query Match 4.6%; Score 69.4; DB 29; Length 494;
Best Local Similarity 48.6%; Pred. No. 2.1e-05;
Matches 190; Conservative 0; Mismatches 201; Indels 0; Gaps 0;
Qy 236 CTGAGTAAAGCTTCAACTGATTTTAACTTTCTCTTTTACGACAACTACTAATGGGGA 295
Db 22 CTGGGGCTTCTACTACTACTAGTACTGCTGCTACTACTACTGCTGCTGCTGCTGCT 81
Qy 296 GAACATATGTAACCTAATAAAGAGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 355
Db 82 CGACTACTGGCAGCAGCTGCTACTACTACTACTGCTGCTGCTGCTGCTGCTACTACTA 141
Qy 356 CTATATCAGTAACTACTACACAGTTTCCACTGTTACTAGTCAAAATAATGATGCTACTG 415
Db 142 CGAGGCTGCTACTACTACTGCTACTACTACTACTACTACTACTACTACTGCTACTACTA 201
Qy 416 AGGTATTAATGTTTCCCAATATGGATATTAATGTTTCTAGCAGCTGTTAATAATGGAG 475
Db 202 CTA 261
Qy 476 CTGTAATTTATGTTTACATCTGTAATCTCAATGCTACTATTTCATGAAAGAACTACTGATG 535
Db 262 CTA 321
Qy 536 CGCATGGGAGATGGTGGTGGAAAACTGTAGATCAAAAAACGTACTCGGTTGGTGATA 595
Db 322 CTACTACTACTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTA 381
Qy 596 CAGTCAATATATCTATTACTATTAGAAATGC 626
Db 382 CTACTACTACTGCTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC 412
RESULT 4
BX173672/c
LOCUS BX173672 829 bp DNA linear GSS 13-MAR-2003
DEFINITION Danio rerio genomic clone DKEY-150M6, genomic survey sequence.
ACCESSION BX173672
VERSION BX173672.1 GI:28005377
KEYWORDS GSS.

SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 829)
AUTHORS Humphray, S.J., Huckle, E. and Durham, J.L.
TITLE Direct Submission
JOURNAL Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humphray@sanger.ac.uk Unpublished
COMMENT This sequence was generated from the SP6 end of BAC 150M6, a part of the Daniokey BAC Library created by R. Plasterk and N.V. Keygene. Further details:
http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES
source 1..829
Location/Qualifiers
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEY-150M6"
/tissue_type="Testis"
/note="vector pindigoBAC-536"
ORIGIN
Query Match 4.6%; Score 69; DB 29; Length 829;
Best Local Similarity 45.2%; Pred. No. 2.7e-05;
Matches 252; Conservative 0; Mismatches 305; Indels 0; Gaps 0;
Qy 385 ACTGTTACTGAGTCAAAATATGATGTTACTGAGCTTATTAACTTTCCCAATATGGATAT 444
Db 645 AATAATAATAATAAGATAATAATAATAATGATGATGATGATGATGATGATGATGATGAT 586
Qy 445 TATTATGTTTCTAGCACTGTTAATAATGGAGCTGTAAATTAATGTTTACATCTGTAACCTCA 504
Db 585 AATAATAGTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 526
Qy 505 AATGCTACTATTTCATGAAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 564
Db 525 AAT 466
Qy 565 GTAGATCAAAAAACGCTACTCGGTTGGTGATACAGTCAAAATATATCTATTACTTATAAGAA 624
Db 465 GATGATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 406
Qy 625 GCAGTCAATTTATCATGCTGACAGAAAAGTGTATCAATATGTTTAAAGGATATGATGCCA 684
Db 405 GATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 346
Qy 685 TCTGCTTCTGCTAGTTGATGAAAGGGTCTTTTGAAGTAACTATTACTGATGGATCA 744
Db 345 AAT 286
Qy 745 GCGAATATTACAACTCTAACTCAAGGTTGCGAAAAAGCACTGGAGATGATAACCTGTTA 804
Db 285 AAT 226
Qy 805 GAGGAAAAATAAATTTTCAGATTACTATTTCGTTGGGCGAGCTACCAATATCTCCCAACCGGA 864
Db 225 AAT 166
Qy 865 AATACTCAAAAATGGAGCTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 924
Db 165 AATAGTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 106
Qy 925 ACTTATACAGGAGTATT 941
Db 105 TAAAAATAGAAAAATATT 89
RESULT 5
CNS02HA4 989 bp DNA linear GSS 01-SEP-2000
LOCUS CNS02HA4

```

DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
139021 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL197365.1 GI:7835515
VERSION AL197365
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodontidae; Tetraodon.
1
REFERENCE 1
AUTHORS Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bertot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE Estimate of human gene number provided by genome-wide analysis
using Tetraodon nigroviridis DNA sequence
JOURNAL Nat. Genet. 25 (2), 235-238 (2000)
MEDLINE 20296633
PUBMED 10835645
REFERENCE 2
AUTHORS Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
Saurin,W., Bertot,A. and Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Genome Res. 10 (7), 939-949 (2000)
MEDLINE 20359837
PUBMED 10899143
REFERENCE 3
AUTHORS (bases 1 to 989)
GENOSCOPE Direct Submission
Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES             source
Location/Qualifiers
1..989
/organism="Tetraodon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone="139021"
/clone_lib="G"
/note="Genoscope sequence ID : COAG139AH11LP1-end : T7"

ORIGIN
Query Match 4.5%; Score 68; DB 29; Length 989;
Best Local Similarity 44.9%; Pred. NO. 4.6e-05;
Matches 213; Conservative 7; Mismatches 254; Indels 0; Gaps 0;

Qy 171 TAATGCAATGTATCTGATTCGAATAAAGATGGAGCTTCATTATTTCCTCAAGTAA 230
Db |||
56 TACVACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC 115
Qy 231 AGAAGCTGAGTATAAAGCTTCAACTGATTTTAAATTCCTTTTACGCACTACTAATGG 290
Db |||
116 TAATACAAAACACTACWACACTACTACTACTACTACTACTACTACTACTACTACTAA 175
Qy 291 AGGAGAACATATGTAATAAAGATACGCTCGTCAGCAAAATGAGATTCGACATGGC 350
Db |||
176 TACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC 235
Qy 351 TAAATCTATATCAGTAAATACTACACCAGTTTCACCTGTTTACTGAGTCAATAATGATGG 410
Db |||
236 TACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTATAC 295
Qy 411 TACTGAGTTATTAAATGTTTCCCAATANGGATATTAATATGTTTCTAGCACTGTTAATAA 470
Db |||
296 TACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC 355

```

```

Qy 471 TGGAGCTGTAATTATGTTACATCTGTAACTCCAAATGCTACTATTTCATGAAAGATAC 530
Db |||
356 TACTGCTAATACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC 415
Qy 531 TGATGCGACATGGGGAGATGGTGTGAAAAAAGCTGTAGATCAAAAAACGTAAGTCTGGTTGG 590
Db |||
416 TACVACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTGGTGG 475
Qy 591 TGATACAGTCAATATATCTACTATTACTTATTAAGATGCGAGTCAATATATCATCGTAC 644
Db |||
476 TGGTCTCTCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTCTCC 529

RESULT 6
AQ946120/c
LOCUS AQ946120.1
DEFINITION Sheared DNA-46J23-TR Sheared DNA Trypanosoma brucei genomic clone
Sheared DNA-46J23, genomic survey sequence.
ACCESSION AQ946120
VERSION AQ946120.1 GI:6769385
KEYWORDS GSS.
SOURCE Trypanosoma brucei
ORGANISM Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
1 (bases 1 to 641)
REFERENCE 1
AUTHORS El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
Gerrard,C., Leech,V., de Jong,P., Ollu,E., Melville,S.,
Donelson,J., Fraser,C. and Adams,M.
TITLE Determination of clone end sequences from Trypanosoma brucei GUTat
10.1 sheared DNA library
JOURNAL Unpublished (1999)
COMMENT Other GSSs: Sheared DNA-46J23.TP
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
DNA library constructed at TIGR. Clones will be available for
distribution through ATCC. Sheared DNA end sequences search page:
http://www.tigr.org/tdb/mdb/tbdb/.
Seq primer: M13-Reverse
Class: shotgun.
FEATURES             Location/Qualifiers
1..641
/organism="Trypanosoma brucei"
/mol_type="genomic DNA"
/strain="TREU927/4 GUTat 10.1"
/db_xref="taxon:5691"
/clone="Sheared DNA-46J23"
/clone_lib="Sheared DNA"
/note="Vector: pUC18; Site 1: SmaI; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
sheared to give a tight size distribution (approx 2 kb).
The v + i method used for the library construction is
described in detail in Smith, H.O. and Venter, J.C.
(Making small insert libraries for whole genome shotgun
sequencing projects. In Genome Sequencing: A Practical
Approach, eds. M. Vaudin and B. Borell, Oxford University
Press, 1999)."
ORIGIN
Query Match 4.5%; Score 67.8; DB 28; Length 641;
Best Local Similarity 45.3%; Pred. NO. 5e-05;
Matches 246; Conservative 0; Mismatches 297; Indels 0; Gaps 0;

Qy 366 TAATACTACACCAGTTTCCTGTTTACTGAGTCAATAATGATGCTACTGAGGTATTAA 425

```


The BACs can be obtained from <http://www.incyte.com>.

```

FEATURES             source
  Location/Qualifiers
    1..501
      /organism="Takifugu rubripes"
      /mol_type="genomic DNA"
      /db_xref="taxon:31033"
      /clone="264E22cA9"
      /clone_lib="BAC 264E22"

ORIGIN
Query Match      4.4%; Score 66.4; DB 29; Length 501;
Best Local Similarity 48.9%; Pred. No. 0.0001;
Matches 178; Conservative 0; Mismatches 186; Indels 0; Gaps 0;

QY 171 TAATGCAATGATATCTGATTCGAATAAGATGGAGCTTCTTATTTAATTCCTCAAGGTAA 230
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 101 TACTACTACTGCTACTACTACTACTACTACTACTACTGCTGCTCTCTACTACTGCTACTAC 160

QY 231 AGAGCTGAGTATTAAGCTTCACGTGATTTTAAATCTCTTTTACGACAACTACTAATGG 290
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 161 TACTACTACTGCTACTACTACTGCTGCTCTCTACTACTACTACTACTACTACTACTAC 220

QY 291 AGGAGAACATATGTAATCAAAAGATACCTGCGTCAGCAATGAGATTGGACATGGC 350
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 221 TAGTACTACTGCTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTGCTGC 280

QY 351 TAAATCTATATCAGCTAATATACACCAAGTTTCCACTGTTTACTGAGTCAAAATATGATGG 410
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 281 TACTACTACTACTGCTACTACTACTACTACTACTACTACTACTGCTCTCTACTACTACTACTAC 340

QY 411 TACTGAGTTTAAATGTTTCCCAATATGGATATTTATGTTTCTAGCAGTGTAAATAA 470
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 341 TACTACTACTACTACTACTGCTGCTTCTACTACTACTACTACTACTACTACTACTACTACTAC 400

QY 471 TGGAGCTGTAATATGTTTACTCTGTAATCTGTAATCCCAATGCTACTATTTCATGAAAGATAC 530
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 401 TACTGCTACTACTACTACTACTACTACTACTACTACTACTACTGCTGCTACTACTACTACTACTAC 460

QY 531 TGAT 534
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 461 TGCT 464

RESULT 9
BX729701
LOCUS
DEFINITION BX729701 XGC-tadpole Silurana tropicalis cDNA clone TTPA075e10 5',
mRNA sequence.
ACCESSION BX729701
KEYWORDS BX729701.1 GI:38402442
SOURCE EST.
ORGANISM Silurana tropicalis (western clawed frog)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
Xenopodinae; Silurana.
1 (bases 1 to 699)
Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (11_2003)
Unpublished (2003)
Contact: Croning MDR
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS SEQUENCE ID: TTPA075e10.p1kSP6
Sequencing primer: SP6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Nigel Garrett.
cDNA was oligo dt primed from Sug of poly A+ RNA from tadpole
embryos. EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end.
Vector: pCS107; Site_1: EcoRI; Site_2: NotI

Host: Escherichia coli DH10B.
Location/Qualifiers
  1..699
    /organism="Silurana tropicalis"
    /mol_type="mRNA"
    /db_xref="taxon:8364"
    /clone="TTPA075e10"
    /dev_stage="tadpole (stage 35-40)"
    /lab_host="E. coli DH10B"
    /clone_lib="XGC-tadpole"
    /notes="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dt primed from Sug of poly A+ RNA from tadpole
embryos. EcoRI-NotI cut cDNA was then ligated into pCS107
with EcoRI at the 5' end and NotI at the 3' end"

ORIGIN
Query Match      4.4%; Score 66.4; DB 13; Length 699;
Best Local Similarity 45.6%; Pred. No. 0.00011;
Matches 272; Conservative 0; Mismatches 321; Indels 3; Gaps 1;

QY 146 ATAAAGTTTTTGATGCGAATAAGATAGTAATGCAATGATCTGATTCGAATAAGATGGAG 205
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 83 ATACTGATCTACTACTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 142

QY 206 CTTCTTATTTAATTCCTCA---AGGTAAAGAGCTGAGTATAAGCTTCAACTGATTTTA 262
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 143 GTTCTGCTACTGCTACTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 202

QY 263 ATTCTCTTTTACGCAACTACTAATGAGGAGGAGAACATATGTAATCAAAAGATAGTCTG 322
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 203 ATACTGCTTCTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 262

QY 323 CGTCGCAATAGATGCGCATGCGCTAAATCTATATCAGCTAATATACACAGTTT 382
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 263 GTACTGATCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 322

QY 383 CCAGCTTCTAGTCAAAATATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 442
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 323 GTACTGATCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 382

QY 443 ATTATTTATGTTTCTAGCAGTCTGTTAATGAGGAGCTGTAATATGTTTCACTGTAACTC 502
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 383 GTACTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 442

QY 503 CAAATGCTACTATTTCATGAAAGAAATACTGATGCGACATGCGGAGATGCTGCTGGAARA 562
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 443 ATACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 502

QY 563 CTGTAGATCAAAAAACGTACTCGGTTGGTGATACAGTCAAAATATATATTTATTAAGA 622
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 503 GTACTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 562

QY 623 ATGCGAGTCAATTTATCATGCTACAGAAAAGTGTATCAATATGTTTAAAGGATACTATGC 682
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 563 ATACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 622

QY 683 CATCTGCTTCTCTAGTTGATTTGAACGAGGCTCTTATGAGTAACTATTACTGAT 738
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 623 GTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 678

RESULT 10
BX780846
LOCUS
DEFINITION BX780846 575 bp DNA linear GSS 14-WAR-2003
genomic clone ii20h11, gi WGS-SpicolorF (DH5A methyl filtered) Sorghum bicolor
genomic survey sequence.
ACCESSION BX780846
VERSION BX780846.1 GI:28958388
KEYWORDS GSS.
SOURCE Sorghum bicolor (sorghum)
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

```

clade; Panicoidae; Andropogoneae; Sorghum.
1 (bases 1 to 575)
REFERENCE Rabinowicz, P.D., O'Shaughnessy, A.L., Balija, V., Dedhia, N.,
AUTHORS Katzenburger, F., King, L., Miller, B., Muller, S., Nascimento, L.,
Zutavern, T., Palmer, L., McCombie, W.R. and Martienssen, R.A.
TITLE Genomic shotgun sequences from Sorghum bicolor (methyl-filtered)
COMMENT Unpublished (2002)
Contact: W. Richard McCombie
Lita Annenberg Hazen Genome Sequencing Center
Cold Spring Harbor Laboratory
PO Box 100, Cold Spring Harbor, NY 11724, USA
Tel: 516 367 8894
Fax: 516 367 8874
Email: mcombie@cshl.org
Plate: 1120 row: h column: 11
Seq primer: -21M13UnivRev
Class: shotgun
High quality sequence stop: 575.
FEATURES
source Location/Qualifiers
1. 575
/organism="Sorghum bicolor"
/mol_type="genomic DNA"
/db_xref="taxon:4558"
/clone="ii20h11"
/lab_host="DH5a"
/clone_lib="WGS-SbicolorF (DH5a methyl filtered)"
/notes="Site 1: Xba I; Site 2: Xba I; The vector was
digested with XbaI and one nucleotide was added by fill in
in the recessive 3' end. The genomic DNA was nebulized,
end repaired, adaptor ligated and size fractionated using
sephadex. The resulting fragments were between 0.8 and 1.9
kb and were cloned into the vector (x/y reads in M13mp19,
b/g reads in pUC19). The same ligation was transformed
into DH5a."

Query Match 4.4%; Score 66.2; DB 28; Length 575;
Best Local Similarity 45.1%; Pred. No. 0.00012;
Matches 245; Conservative 0; Mismatches 238; Indels 0; Gaps 0;
QY 137 ATAAGCATATAAGTTTTTGATGACAGAAATAGATAATGCAATGTATCTGATTCGAATA 196
DB 3 ATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATACTAGTA 62
QY 197 AAGATGGAGCTTCCTATTATTCCTCAAGGTAAAGAGCTGAGTATATAAGCTTCAACTG 256
DB 63 GTAGTAGTAGTAATAATAAGTAGTAATAATAATACTAATAATAATAATAATAAGCA 122
QY 257 ATTTTAAATCTCTTTTACAGCAACTACTAATGAGGGAGCAACATATGTAACTAAAAAG 316
DB 123 ATAATAATAATAATGACTAATAATAATAATAAGTAGTAATAATAATAATAATAATA 182
QY 317 ATACTGGTCAGCAAAATGAGATGCGACATGGGCTAAATCTATATCAGCTAAATACTACAC 376
DB 183 ATATACTAATAATAATACTAATAATAATAATAATAATAATAATAATAATAATAATACTA 242
QY 377 CAGTTTCCACTGTTTACTGAGTCAAAATGAATGATGCTAGGTTTATTAATGTTTCCCAAT 436
DB 243 ATATAATACTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 302
QY 437 ATGATATTAATATGTTTCTAGCACTGTTAATAATGAGGCTGTAATTTATGGTTACATCTG 496
DB 303 ATATAATACTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 362
QY 497 TAATCCAAATGCTACTATTTCATGABAAGATCTGATGCGACATGGGAGAGTGGTG 556
DB 363 ATATACTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 422
QY 557 GAAAACTGTAGATCAAAAAACGCTACTCGGTGGTGAACAGTCAAAATATACTATTACTT 616
DB 423 CTAATAGTAGTAATACTAATGAATAATAATAATAATAATAATAATAATAATAATAATA 482
QY 617 ATAAGATGCAAGTCAATTATCATGGTACAGAAAAAGGTATCAATATGTTTATAAAGGATA 676

Db 483 GTATAATAGTAGTAATAATAAGTAGTAGTAATAATAATAATAATAATAATAATAATAATA 542
QY 677 CTA 679
Db 543 CTA 545
RESULT 11
CNS012TP 1101 bp DNA linear GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC
DEFINITION BACN08A01 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL102007.1 GI:5613618
VERSION ALL102007
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Bukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segret@genoscope.cns.fr)
COMMENT - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
Project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelOBAC11.
FEATURES
source Location/Qualifiers
1. 1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN08A01"
/clone_lib="drosBAC"
/plasmid="pBelOBAC11"
/notes="end : T7"
ORIGIN
Query Match 4.4%; Score 66.2; DB 29; Length 1101;
Best Local Similarity 30.4%; Pred. No. 0.00012;
Matches 269; Conservative 131; Mismatches 486; Indels 0; Gaps 0;
QY 330 AAATGAGATTGCGACATGGGCTAAATCTATATCAGCTAATACTACACAGTTTCCACTGT 389
DB 1033 AWWKSWNWGWAAAAGGVTGAWAAGHTGSWABKTTTGTGKWKCKTWTATSKWBT 974
QY 390 TACTGAGTCAAAATPAATGATGGTACTGAGGTATTAATGTTTCCCAATATGATATTAATA 449
DB 973 AWTAAARWAWDAAGNBTTCKAKWTTKWTMTAWMTKAWTKWTTWAAATATKRTTBWBTAT 914
QY 450 TGTTCCTAGCACTGTTAATAATAATGAGCTGTAATATGTTTACATCTGTAATCTCAATGC 509
DB 913 TATTGTTGTMYPABTATWTGHCRAWTTMWAATASMYAAAATVTAATAAKATRAGMGA 854
QY 510 TACTATTTCATGAAAGAAATACTGATCGACATGGGAGATGGTGGTGGAAAAAAGCTGAGA 569
DB 853 TMCMAKAAKAWBATANAWKGTCTGKGBWMMWTKTTTXYTASCTWTMTATMYAKA 794
QY 570 TCAAAAAACGCTACTCGGTTGGTGATACAGTCAAAATATATCTATTATTAAGAATGAGT 629
DB 793 MGTWTMTCKAATWAGTTMTACMAYNGTAYTWCTTACMTATTTTGTGKAWTTAAGGAHA 734
QY 630 CAATTATCATGGTACAGAAAAAGTGTATCAATATGTTTATAAAGGATACTATGCCATCGC 689

SOURCE Tetraodon nigroviridis
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontoidea; Tetraodontidae; Tetraodon.
1
REFERENCE Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
AUTHORS Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W., and Weissenbach,J.
TITLE Estimate of human gene number provided by genome-wide analysis
using Tetraodon nigroviridis DNA sequence
JOURNAL Nat. Genet. 25 (2), 235-238 (2000)
MEDLINE 20296633
PUBMED 10835645
2
REFERENCE Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
AUTHORS Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
Saurin,W., Bernot,A. and Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Genome Res. 10 (7), 939-949 (2000)
MEDLINE 10899143
PUBMED 10899143
REFERENCE 3 (bases 1 to 735)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES Location/Qualifiers
source
1..735
/organism="Tetraodon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone="123M05"
/clone_lib="G"
/note="Genoscope sequence ID : COBG123AG03LP1-end : T7"
ORIGIN
Query Match 4.3%; Score 64.6; DB 29; Length 735;
Best Local Similarity 44.7%; Pred. No. 0.00027;
Matches 193; Conservative 9; Mismatches 230; Indels 0; Gaps 0;
Qy 215 TAATTCCTCAAGGTAAGAGCTGAGTATAAGCTTCAAGCTTTAAATTCCTTTTA 274
Db 597 TRCWRCTACWACTAANNAANGCNGCGNACNACTACTAATAACAACACTACTGCTA 538
Qy 275 CGACAACTACTAATGAGGAGGAGACATATGTAATAAAAGATACCTCGCTCAGCAATG 334
Db 537 CTACTACTACTGTAAGAACTACTACTGCTACTGCTGCTCTACTACTGCTACTGCTACWA 478
Qy 335 AGATTGCGACATGGCTAAATCTATATCAGCTAATACACAGTTTCCACTGTTACTG 394
Db 477 TTKWGTGCTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTGCTGCTG 418
Qy 395 AGTCAATAATGATGGTACTGAGTTATTAATGTTCCCAATATGATATATATGTTT 454
Db 417 CTACTACTGCTACTACTACTGCTGCTACTACTGCTGCTGCTGCTGCTACTA 358
Qy 455 CTAGCACTGTTAATAAGGAGCTGTAATATGTTTACATCTGTAATCCAAATGCTACTA 514
Db 357 CTACTACTACTACTACTGCTGCTACTACTACTACTACTACTACTACTACTACTACTA 298
Qy 515 TTCATGAAAAGAACTACTGATCGGACATGGGAGATGGTGGGAAAACACTGTAGACAAA 574
Db 297 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 238
Qy 575 AAAGCTACTCGTGGTGATACAGTCAAAATATACTATTACTTATAAGAAATGCAGTCAAT 634

Db 237 CTACTGCTGCTGCTGCTACTACTACTACTACTACTACTACTACTACTACTACTA 178
Qy 635 ATCATGTTACAG 646
Db 177 CTACTGCTACTG 166
RESULT 14
BZ643398
LOCUS OGAOW577C ZM 0.7 1.5 KB Zea mays genomic clone ZMMBMA011J17,
DEFINITION genomic survey sequence.
ACCESSION BZ643398
VERSION BZ643398.1 GI:28104877
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 480)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
TITLE Consortium for Maize Genomics
JOURNAL Unpublished (2002)
COMMENT Other GSSs: OGAOW57TM
Contact: Cathy Whitelaw
TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
FEATURES Location/Qualifiers
source
1..480
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMA011J17"
/clone_lib="ZM 0.7 1.5 KB"
/note="Vector: pBSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"
ORIGIN
Query Match 4.2%; Score 62.8; DB 28; Length 480;
Best Local Similarity 48.8%; Pred. No. 0.0007;
Matches 200; Conservative 0; Mismatches 207; Indels 3; Gaps 1;
Qy 130 GCAACCTATAAAGCATATAAAGTTTTTGTATGCAAAATAGATAATGCAATGTATCTGAT 189
Db 17 GCTACTACTACTGCTACTACTGCTGCTGCTGCTGCTGCTGCTGCTACTGCTACTGCT 76
Qy 190 TCGAATAAAGATGGAGCTTCTTATTTAATTCCTCAAG---GTAAAGAGCTGAGTATAAA 246
Db 77 ACTACTACTACTGCTGCTGCTACTACTACTACTGCTAATGCTACTACTACTGCTGCTACT 136
Qy 247 GCTTCAACTGATTTAATTCCTCTTTTACGACAACACTACTAATGGAGGAGCAATATGTA 306
Db 137 ACTACTACTACTGCTGCTGCTACTACTGCTGCTGCTGCTGCTGCTGCTACTACTGCT 196
Qy 307 ACTAAAAAGATAGCTCGTCAGCAAAATGAGATTGCCACATGGGCTAAATCTATATCAGCT 366
Db 197 ACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTGCT 256
Qy 367 AATPACTACACCGTTTCCACTGTTACTGAGTCAAAATAGATGCTACTGAGTTATTAAT 426
Db 257 ACTACTGCTACTACTACTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTACT 316
Qy 427 GTTTCCCAATATGATATTAATTTCTAGCACTGTTAATTAATGAGCTGTAATATG 486

Db 317 GCTACTACTGCTACTGCTGCTACTGCTACTGCTACTGCTACTGCTACTGCT 376

Qy 487 GTTACATCTGTAACCTCAAATGCTACTATTTCATGAAAGAAATACTCATGC 536

Db 377 ACTGCTGCTACTACTGCTACTGCTACTGCTAACTGCTACTACTGCTACTACTAC 426

RESULT 15					
BZ423698					
LOCUS	BZ423698	513 bp	DNA	linear	GSS 10-DEC-2002
DEFINITION	id52c10.g1 WGS-SbicolorF (DH5a methyl filtered) Sorghum bicolor genomic clone id52c10 5', genomic survey sequence.				

ACCESSION	B2423698	
VERSION	B2423698.1	GI:26373196
KEYWORDS	SS.	
SOURCE	Sorghum bicolor	(sorghum)
ORGANISM	Sorghum bicolor	
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.	

REFERENCE
AUTHORS
1 (bases 1 to 513)
Rabinovich, P.D., O'Shaughnessy, A.L., Ballja, V., Dedhia, N.,
Katzenburger, F., King, L., Miller, B., Muller, S., Nascimento, L.,
Zutavern, T., Palmer, L., McCombie, W.R. and Martienssen, R.A.
TITLE
JOURNAL
Genomic shotgun sequences from Sorghum bicolor (methyl-filtered)
Unpublished (2002)

COMMENT:
Captured on 1/10/02,
Contact: W. Richard McCombie
Lita Annenberg Hazen Genome Sequencing Center
Cold Spring Harbor Laboratory
PO Box 100, Cold Spring Harbor, NY 11724, USA
Tel: 516 367 8884

Tel: 516 367 8884
 Fax: 516 367 8874
 Email: mccombie@cshl.org
 Plate: id52 row: c column: 10
 Seq primer: -21M13univRev
 Class: shotgun
 High quality sequence stop: 513

```

FEATURES
    source
        high quality sequence sc0p1: 513.
        Location/Qualifiers
            1..513
                /organism="Sorghum bicolor"
                /mol_type="genomic DNA"
                /db_xref="taxon:4558"
                /clone="id52c10"
                /lab_host="DH5a"
                /clone_lib="WGS-SbicolorP (DH5a methyl filtered)"
                /note="Site 1: Xba I; Site 2: Xba I; The vector was
                    digested with XbaI and one nucleotide was added by fill in
                    in the recessive 3' end. The genomic DNA was nebulized,
                    end repaired, adaptor ligated and size fractionated using
                    sephadex. The resulting fragments were between 0.8 and 3
                    kb and were cloned into the vector (.x/y reads in M13mp19,
                    .b/g reads in pUC19). The same ligation was transformed
                    into DH5a."

```

ORIGIN

Query Match	4.1%	Score	62.6;	DB	28;	Length	513;	
Best Local Similarity	46.3%;	Pred. No.	0.00078;					
Matches	206;	Conservative	0;	Mismatches	239;	Indels	0; Gaps	0;

171	TAATGCAAAAGTATCTCATTTCCGAATAAAGATGGACGTTCTTTATTTCCTCAAGTAA	230
Ddb		
65	TACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT	124
231	AGAACTTGAGTATAAAAGCTTCAAAGTATTTAAATTTCTTTTTACGACAACACTACTAATGG	290
Ddb		
125	TACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT	184
291	AGGAGAGAACATANGTACTATAAAAAGATCTCGGTGAGCAAAATGAGATTCGCACATGGGC	350
Ddb		
185	TAGTGTCATCTACTACTACTACTACTCTGCTACTACTACTACTACTACTACTACTACTACT	244
351	TAAATCTATATCAGCTAANTACTACACCAGTTTCCAAGTGTCTTACTGAGTCAAAATATGATGG	410
Ddb		

[illegible]

Search completed: September 13, 2004, 04:42:33
Job time : 4106.31 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2004, 22:30:05 ; Search time 111.254 Seconds
(without alignments)
7527.109 Million cell updates/sec

Title: US-10-009-254-1

Perfect score: 1509

Sequence: 1 atgaaagaaatgattca.....gtcgtgcttaccgtttcttaa 1509

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*

2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*

3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*

4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*

5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*

6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	64.8	4.3	1980	4	US-09-134-000C-2719
2	64.8	4.3	1983	4	US-09-071-035-217
3	56.2	3.7	1575	4	US-09-543-581A-615
4	50.4	3.3	14066	4	US-09-601-198-56
5	49.8	3.3	1687	4	US-09-071-035-219
6	48.4	3.2	3666	2	US-08-682-517-13
7	48.4	3.2	3666	2	US-08-682-517-14
8	48.4	3.2	4197	2	US-08-682-517-7
9	48.4	3.2	4197	2	US-08-682-517-8
10	47.8	3.2	740	3	US-09-451-117-1
11	47.8	3.2	740	4	US-09-888-655-1
12	47.4	3.1	1037	4	US-09-181-585-3
13	47.4	3.1	1159	4	US-09-181-585-1
14	47.4	3.1	1471	4	US-09-181-585-2
15	47.4	3.1	3945	4	US-09-200-6508-6
16	47.4	3.1	4358	4	US-08-956-171E-454
17	47.2	3.1	7218	1	US-08-232-463-14
18	46.8	3.1	1948	1	US-07-849-438-1
19	46.8	3.1	4590	4	US-09-134-001C-1108
20	45.6	3.0	4376	1	US-08-119-125A-1
21	45.6	3.0	6744	1	US-08-119-125A-2
22	45	3.0	832	4	US-09-621-976-2813
23	44.6	3.0	11679	4	US-09-328-352-1377
24	44.2	2.9	606	4	US-09-601-198-166
25	44.2	2.9	30549	4	US-09-134-001C-322
26	44.2	2.9	1716	4	US-09-134-001C-1028
27	43.4	2.9	2763	4	US-09-463-402-5

28	43.4	2.9	2766	4	US-09-889-572-13	Sequence 3, Appli
29	43.4	2.9	4085	4	US-09-463-402-13	Sequence 13, Appl
30	43.2	2.9	5181	1	US-08-257-073-10	Sequence 10, Appl
31	43	2.8	2001	4	US-09-528-784A-84	Sequence 84, Appl
32	43	2.8	2001	4	US-09-569-098A-84	Sequence 84, Appl
33	43	2.8	3057	4	US-09-601-198-55	Sequence 55, Appl
34	43	2.8	3402	4	US-09-528-784A-86	Sequence 86, Appl
35	43	2.8	3402	4	US-09-569-098A-86	Sequence 86, Appl
36	42.8	2.8	986	4	US-09-134-001C-1789	Sequence 1789, Ap
37	42.6	2.8	168575	4	US-09-426-290-1	Sequence 1, Appli
38	42.2	2.8	810	4	US-09-601-198-31	Sequence 31, Appl
39	42.2	2.8	3292	1	US-07-814-964-12	Sequence 12, Appl
40	42.2	2.8	3292	1	US-08-258-442-12	Sequence 12, Appl
41	42.2	2.8	3292	1	US-08-328-809-7	Sequence 7, Appli
42	42.2	2.8	3292	4	US-08-866-840-7	Sequence 7, Appli
43	42.2	2.8	3292	5	PCT-US92-11107-12	Sequence 12, Appl
44	42.2	2.8	4185	4	US-09-417-485D-7	Sequence 7, Appli
45	42.2	2.8	10640	4	US-09-417-485D-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1

US-09-134-000C-2719

; Sequence 2719, Application US/09134000C

; Patent No. 6617156

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

; FILE REFERENCE: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS

; CURRENT APPLICATION NUMBER: US/09/134,000C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/055,778

; PRIOR FILING DATE: 1997-08-15

; NUMBER OF SEQ ID NOS: 6812

; SOFTWARE: Patent in version 3.1

; SEQ ID NO 2719

; LENGTH: 1980

; TYPE: DNA

; ORGANISM: Enterococcus faecalis

; US-09-134-000C-2719

Query Match 4.3%; Score 64.8; DB 4; Length 1980;
Best Local Similarity 52.1%; Pred. No. 9.4e-07;
Matches 172; Conservative 0; Mismatches 152; Indels 6; Gaps 1;

Qy	1180	ACAGAAGCTTAATCGAACAGATATACACAGGAGCAGATGGTATATACCATTCACAGGC	1239
Db	1657	ACAAAAGCTGAACCAACTACTTTTACCAACACCGCTGATGGATTAGTTGATATCACAGGG	1716
Qy	1240	TTCAAGAAGGTACATACTATCTAGTTGAGAAAAGGCTCCCTTAGGTTCATTTGTTA	1299
Db	1717	CTTAATACGGTACCTTATTTAGAGAACTGTAGTCTCTGATGATTATGTTGTTA	1776
Qy	1300	GATAACTCTCAGAAGGTTATTTTAGGAGATGGAGCCACTGATACGACTAATTCAGATAAC	1359
Db	1777	ACAAATCGGATTGAATTTGGTCAATGAACAATCATATGCGACAA-----CAGAAAAC	1830
Qy	1360	CTTTTAGTTAACCAACTGTTGAAATAACAAAGGTACTAGTTGCTTCAACAGGTGGT	1419
Db	1831	CTAGTTTTCACGAAAAAGTACCAACAAACACAAAGGTACCTTACCTTCAACAGGTGGC	1890
Qy	1420	ATTGGTACCACAATTTTCTACATTATAGTGGTCAATTTTAGTATAGGAGCAGGTATCGTG	1479
Db	1891	AAAGGAATCTACGTTTACITTAGGAAGTGGCGAGCTTGTCTACTATTGAGGAGTCTAC	1950
Qy	1480	CTTGTGCTCGTCGTCGTTTACGTTCTTAA	1509
Db	1951	TTTGCTAGACGTAGAAAAGAAATGCTTAA	1980

RESULT 2
US-09-071-035-217
; Sequence 217, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 217:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1983 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-09-071-035-217

Query Match 4.3%; Score 64.8; DB 4; Length 1983;
Best Local Similarity 52.1%; Pred. No. 9.4e-07;
Matches 172; Conservative 0; Mismatches 152; Indels 6; Gaps 1;

QY 1180 ACAGAGCTAATGCAACAGAAATATACAAAGGAGCAGATGATTAATACCATTTACAGGC 1239
DB 1660 ACAAGAGCTGAGCAACTACTTTTACAAACGCGTGATGATTTGATATCACAGGG 1719

QY 1240 TTGAAAGAGGTACATATCTAGTTGAGAAAAGGCTCCCTTAGGTTACAAATTTGTTA 1299
DB 1720 CTTAAATACGGTACCTATTATTAGAGAAAGCTGCTGATGATTTGTTA 1779

QY 1300 GATACTCTCAGAGGTTATTTAGGAGATGGAGCCACTGATAGCACTAATTCAGATAAC 1359
DB 1780 ACAATCGGATGAATTTGTTGGTCAATGAACAATCATATGGCCAA-----CAGAAAAC 1833

QY 1360 CTTTTAGTTAAACCAACTGTTGAAAATAACAAAGGTACTCAGTTGCTTCCAAACAGGTGGT 1419
DB 1834 CTAGTTTACCAGAAAAGTACCACAAACACAAAGGTACCTTACCTTCACAGGTGGC 1893

QY 1420 ATTGGTACAAATTTTCTACATTTATAGGTGCAATTTTAGTAATAGGAGCAGGTATCGTG 1479
DB 1894 AAAGGAATCTACGTTTACTTAGGAAGTGGCGAGCTTTGCTACTTTATGCGAGGACTAC 1953

QY 1480 CTTGTTGCTCGTCTGCTTTACGTTCTTAA 1509
DB 1954 TTTGCTAGAGGTAGAAAAGAAATGCTTAA 1983

RESULT 3
US-09-543-681A-615

; Sequence 615, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 615
; LENGTH: 1575
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-615

Query Match 3.7%; Score 56.2; DB 4; Length 1575;
Best Local Similarity 43.3%; Pred. No. 0.00012;
Matches 262; Conservative 0; Mismatches 343; Indels 0; Gaps 0;

QY 107 CAGTTCAAGATACTCAAAAAGCGCGCAACCTATAAAGCATATATAAGTTTGTGATGAGAAA 166
DB 404 CAATTTTTTCTAATGAAATTCGCAATGAAATGCTATATGAAAAGTTAGATAAAAAATGAGC 463

QY 167 TAGATAATGCAATGTATCTGATTTCGAATAAAGATGGAGCTTCTTATTTAATTCCTCAAG 226
DB 464 TTAATGATCAAAATTTTTCGCTGAGTTGATTTGTAATGAAATAGATAATAAATACGTTCTGGTG 523

QY 227 GTAAAGAAGCTGAGTATAAAGCTTCAACTGATTTTAAATCTCTTTTTCGACAACTACTA 286
DB 524 AAATAGACTTAAAGCTTAAATAAGTAACCTTAAATCAAGTTAACTTTAGTGCATAAGAC 583

QY 287 ATGGAGGAGACATATGTAACTTAAAGAGATGCTCGGTGAGCAATGAGATTGCGACAT 346
DB 584 AGCTCGTAAGCAAGCGCATAGAAAATGATAACTTAACCTTAGATACAGAAAATGATAATA 643

QY 347 GGGCTAAATCTATATCAGCTTAATACTACACCACTTTCCACTGTTACTGAGTCAAAATAATG 406
DB 644 TTAACCTTAGAAAATGATAATATTAACTTAGAAAATGATAATATTAACTTAGAAAATGATA 703

QY 407 ATGGTACTGAGGTATTAAATGTTTCCCAATATGGAATATTATTATTTCTAGCACTGTTA 466
DB 704 ATATTAACTTAGAAAATGATAATATTAACTTAGAAAATGATAATATTAACTTAGAAAATG 763

QY 467 ATAAATCGAGCTGTTAATTTGTTTACATCTGTAACCTCCAAATGCTACTATTTCATGAAAAGA 526
DB 764 ATAATATTAACTTAGAAAATGATAATATTAACTTAGAAAATGATAATATTAACTTAGAAA 823

QY 527 ATACTGATCGACATGGGAGATGGTGGTGGAAAACCTGATAGATCAAAAACCGTACTCGG 586
DB 824 ATGATAATGTTTAACTTAGAAAATGATAATGTTTAACTTAGAAAATGATAATGTTTAACTTAG 883

QY 587 TTGGTGATACAGTCAAAATATATTTTACTTTAAGAAATGCAAGTCAATTTATGTTGTTACAG 646
DB 884 AAAATGATAATGTTTAACTTAGAAAATGATAATATTAACTTAGAAAATGATAATGTTTAACT 943

QY 647 AAAAGGTGATCAATATGTTTATAAGGATGACTATGCCACTCTGCTTCTGTTAGTTGATTGA 706
DB 944 TAAATATAAAAAGGCACTTAATAAAGATATAAATAGAAAATACTTATATTTTCAATGGTA 1003

QY 707 ACGAA 711
DB 1004 AAGAA 1008

RESULT 4
US-09-601-198-56/c
; Sequence 56, Application US/09601198
; Patent No. 6531583
; GENERAL INFORMATION:
; APPLICANT: Cassell, Gail H.

Db 4567 ATATAATAATA 4556

|||||

RESULT 5

US-09-071-035-219

; Sequence 219, Application US/09071035

; Patent No. 6448043

; GENERAL INFORMATION:

; APPLICANT: Gil H. Choi

; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides

; NUMBER OF SEQUENCES: 496

; CORRESPONDENCE ADDRESS:

; ADDRESSER: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

```

; COMPIER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 219:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1687 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-09-071-035-219

Query Match 3.3%; Score 49.8; DB 4; Length 1687;
Best Local Similarity 62.4%; Pred. NO. 0.0047;
Matches 78; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 1180 ACAGAAGCTAATGCAACAGATATATACACAGGACGAGATGGTATTAATTTACCATTCACAGC 1239
Db 1466 ACAAAAGCTGAAGCAACTACTTTTACAAACGGCTGATGGATTAGTTGATATCACAGG 1525

Qy 1240 TTGAAGAGGTTACATACTATCTAGTTCGAAAAAAGGCTCCCTTAGGTTACAATTTGTTA 1299
Db 1526 CTTAAATAGGTCACCTATTATTTAGAGAAACTGTAGCTCCTGATGATTAATGTCCTTGTTA 1585

Qy 1300 GATAA 1304
Db 1586 ACAAA 1590

RESULT 6
US-08-682-517-13
; Sequence 13, Application US/08682517
; Patent No. 5874267
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

[illegible]

```

RESULT 8
US-08-682-517-7
; Sequence 7, Application US/08682517
; Patent No. 5874267
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/682,517
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4197 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-682-517-7
Query Match 3.2%: Score 48.4: DB 2: Length 4197

```

Best Local Similarity 44.5%; Pred. No. 0.014; DB 2; Length 4197;
Matches 193; Conservative 0; Mismatches 241; Indels 0; Gaps 0;

QY 302 ATGTAACATAAAGAAAGTACTGCGTCAGCAAAATGAGATTGGACATGGGCTAAATCTATAT 361
|||||
Db 2952 ATGTAAGTGAACAACACTGTTGATGACTGCAACTGTTTCATTAAAGATAGTGCAATAATT 3011
|||||
QY 362 CAGCTAATACTACACCAAGTTTCCACTGTTTACTGAGTCAAAATAAGATGGTACTGAGTTA 421
|||||
Db 3012 CATTATCTCTTACATTAGTTTGAAGTGGTGTAAATACAGGTGATTGCTACCAACTGTTTC 3071
|||||
QY 422 TTAATGTTTCCCAATATGATATATATATGTTTCTAGCACTGTTAATAATGGAGCTGTAA 481
|||||
Db 3072 AAGCTGGTACATTTACTTCTTAACTGCTGGTACATTTAAACAGTTACTTATGCGAGATGCTA 3131
|||||
QY 482 TTATGTTTACATCTGTAACCTCCAAATGCTACTATTTCATGAAAAAGAAATACATGATCGGACAT 541
|||||
Db 3132 AAAATGCTGAGGTGTTGCTGAAAAATATTACTGCTAGCTAACTAAAGAAAACTACTG 3191
|||||
QY 542 GGGAGATGTTGGTGGAAAACTGTAGATCAAAAAAGTACTCGGTTGGTATACAGTCA 601
|||||
Db 3192 GAGCAATTTACTTCTGATACATTTTACACAAGGTGTTTACCATCAGCAGCTACAGCAGCTG 3251
|||||
QY 602 AATATACTATTACTTATAGAAATGCGAGTCAATTTATCATGTCACAGAAAAAGTGTATCAAT 661
|||||
Db 3252 AATATACTTCTTAAATCAATTTGCTGCGAGTTATATACATTTGCAACAGGTGAAGGATTCATT 3311
|||||
QY 662 ATGTTATAAGGATGATGCTATGCTCTTCTGTTAGTTGATTGAAACGAAGGGTCTTTATG 721
|||||
Db 3312 TAAATATTGATATGCTGGTGTCTCAAGTAATTAACCTAGCAGGTAAAAAGGTGCACAAG 3371
|||||
QY 722 AAGTAATATTACT 735
|||||
Db 3372 GTGTAGCTGATGCT 3385

RESULT 9
US-08-682-517-8
; Sequence 8, Application US/08682517
; Patent No. 5874267
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (BPO)
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4197 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Bacillus sphaericus
; INDIVIDUAL ISOLATE: P-1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 95..3850
; FEATURE:
; NAME/KEY: mat peptide
; LOCATION: 185..3850
; FEATURE:
; NAME/KEY: sig peptide
; LOCATION: 95..184
; US-08-682-517-8

Query Match 3.2%; Score 48.4; DB 2; Length 4197;
Best Local Similarity 44.5%; Pred. No. 0.014;
Matches 193; Conservative 0; Mismatches 241; Indels 0; Gaps 0;

QY 302 ATGTAACATAAAGAAAGTACTGCGTCAGCAAAATGAGATTGGACATGGGCTAAATCTATAT 361
|||||
Db 2952 ATGTAAGTGAACAACACTGTTGATGACTGCAACTGTTTCATTAAAGATAGTGCAATAATT 3011
|||||
QY 362 CAGCTAATACTACACCAAGTTTCCACTGTTTACTGAGTCAAAATAAGATGGTACTGAGTTA 421
|||||
Db 3012 CATTATCTCTTACATTAGTTTGAAGTGGTGTAAATACAGGTGATTGCTACCAACTGTTTC 3071
|||||
QY 422 TTAATGTTTCCCAATATGATATATATATGTTTCTAGCACTGTTAATAATGGAGCTGTAA 481
|||||
Db 3072 AAGCTGGTACATTTACTTCTTAACTGCTGGTACATTTAAACAGTTACTTATGCGAGATGCTA 3131
|||||
QY 482 TTATGTTTACATCTGTAACCTCCAAATGCTACTATTTCATGAAAAAGAAATACATGATCGGACAT 541
|||||
Db 3132 AAAATGCTGAGGTGTTGCTGAAAAATATTACTGCTAGCTAACTAAAGAAAACTACTG 3191
|||||
QY 542 GGGAGATGTTGGTGGAAAACTGTAGATCAAAAAAGTACTCGGTTGGTATACAGTCA 601
|||||
Db 3192 GAGCAATTTACTTCTGATACATTTTACACAAGGTGTTTACCATCAGCAGCTACAGCAGCTG 3251
|||||
QY 602 AATATACTATTACTTATAGAAATGCGAGTCAATTTATCATGTCACAGAAAAAGTGTATCAAT 661
|||||
Db 3252 AATATACTTCTTAAATCAATTTGCTGCGAGTTATATACATTTGCAACAGGTGAAGGATTCATT 3311
|||||
QY 662 ATGTTATAAGGATGATGCTATGCTCTTCTGTTAGTTGATTGAAACGAAGGGTCTTTATG 721
|||||
Db 3312 TAAATATTGATATGCTGGTGTCTCAAGTAATTAACCTAGCAGGTAAAAAGGTGCACAAG 3371
|||||
QY 722 AAGTAATATTACT 735
|||||
Db 3372 GTGTAGCTGATGCT 3385

RESULT 10
US-09-451-117-1
; Sequence 1, Application US/09451117
; Patent No. 6277973
; GENERAL INFORMATION:
; APPLICANT: Jenkins, Mark C.
; APPLICANT: Fayer, Ronald
; APPLICANT: Trout, James
; TITLE OF INVENTION: Cloning and Expression of a DNA Sequence Encoding A 41
; Patent No. 6277973
; TITLE OF INVENTION: kDa Cryptosporidium parvum Oocyst Wall Protein
; FILE REFERENCE: 0046.99
; CURRENT APPLICATION NUMBER: US/09/451,117
; CURRENT FILING DATE: 1999-11-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 740
; TYPE: DNA
; ORGANISM: Cryptosporidium parvum
; US-09-451-117-1

Query Match 3.2%; Score 47.8; DB 3; Length 740;
Best Local Similarity 46.7%; Pred. No. 0.011;
Matches 151; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

QY 217 ATTCCTCAAGGTAAGAAGCTGAGTATAAAGCTTCAACTGATTTTAATTTCTTTTACG 276
|||||
Db 370 ACTACTCAATCAATGGAAGTACTGATATAATTTGGTAAATATGATGCTGTTATTATACA 429
|||||
QY 277 ACAACTACTAATGGAGGAGACATATGTAACTAAAAAGATCTGCGTCAGCAAAATAG 336
|||||
Db 430 TCTAGTGGTATCTTAGCTGTTACTTAATACTTAATCTTAATGTTAAATAATAGTAATAGT 489
|||||
QY 337 ATTGGCAGATGGGCTAAATCTATATACAGTAACTACTACACAGTTTCCACTGTTACTGAG 396
|||||

Db 490 AATATTGGATCAGGAATTTTATACAGTTGGTACTTGTTCTTCTACTAGTATTGGTAAT 549
QY 397 TCAAAATAATGATGGTACTGAGGTATTAAATGTTTCCCAATATGGATATATATGTTTCT 456
Db 550 AGTAATGGTGTGCTTTTACTGCTATTTCATCCTAATAATAACAATAGCAATAATTAAT 609
QY 457 AGCACTGTTAATAATGAGGCTGTAATTATGTTTACATCTGTAACCTCCAAATGCTACTATT 516
Db 610 AATAATAATAATAATAAGTAATACCACTCTTACTACTGTTGCTACTAATGCTAATATT 669
QY 517 CATGAAAGAATACTGATGGAC 539
Db 670 ACTACTAATACTACTAATACTACTAC 692

RESULT 11
US-09-888-655-1
; Sequence 1, Application US/09888655
; Patent No. 6521229
; GENERAL INFORMATION:
; APPLICANT: Jenkins, Mark C.
; APPLICANT: Fayer, Ronald
; APPLICANT: Trout, James
; TITLE OF INVENTION: Cloning and Expression of a DNA Sequence Encoding A 41
; Patent No. 6521229
; TITLE OF INVENTION: kDa Cryptosporidium parvum Oocyst Wall Protein
; FILE REFERENCE: 0046.99
; CURRENT APPLICATION NUMBER: US/09/888,655
; CURRENT FILING DATE: 2001-06-26
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 740
; TYPE: DNA
; ORGANISM: Cryptosporidium parvum
US-09-888-655-1

Query Match 3.2%; Score 47.8; DB 4; Length 740;
Best Local Similarity 46.7%; Pred. No. 0.011;
Matches 151; Conservative 0; Mismatches 172; Indels 0; Gaps 0;
QY 217 ATTCTCAAGTAAAGAGCTGAGTATAAAGCTTCAACTGATTTTAAATTTCTTTTACG 276
Db 370 ACTACTCAATCAATGGAAGTTTACTGATCAATAATGTTGGTAATGAGTGTATATTACA 429
QY 277 ACACTACTAATGAGGGAGAACATATGTAATAAAAAAGACTGCGTCAGCAAAATGAG 336
Db 430 TCTAGTGTGATTCTATAGCTGTTACTAATAATCTTAATGGTAATAATAATAGTATAGT 489
QY 337 ATTGGACATGGGCTAAATCTATATCAGCTAATACTACACAGTTTCCACTGTACTGAG 396
Db 490 AATAATGGATCAGGAATTTTATACCAAGTTGGTACTTGTCTTCTACTAGTATTGGTAAT 549
QY 397 TCAATAATGATGTTACTGAGGTATTAAATGTTTCCCAATATGATATATTATGTTTCT 456
Db 550 AGTAATGGTGTGCTTTTACTGCTATTTCATCCTAATAATAACAATAGCAATAATTAAT 609
QY 457 AGCACTGTTAATAATGAGGCTGTAATTATGTTTACATCTGTAACCTCCAAATGCTACTATT 516
Db 610 AATAATAATAATAATAAGTAATACCACTCTTACTACTGTTGCTACTAATGCTAATATT 669
QY 517 CATGAAAGAATACTGATGGAC 539
Db 670 ACTACTAATACTACTAATACTACTAC 692

RESULT 12
US-09-181-585-3
; Sequence 3, Application US/09181585
; Patent No. 6524791
; GENERAL INFORMATION:
; APPLICANT: Ranum, Laura P.W.

; APPLICANT: Koob, Michael
; TITLE OF INVENTION: SPINOCEREBELLAR ATAXIA TYPE 8 AND METHODS OF DETECTION
; FILE REFERENCE: 1100900101
; CURRENT APPLICATION NUMBER: US/09/181,585
; CURRENT FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1037
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: cDNA
; OTHER INFORMATION: comprising exons E, C, and A
US-09-181-585-3

Query Match 3.1%; Score 47.4; DB 4; Length 1037;
Best Local Similarity 47.2%; Pred. No. 0.016;
Matches 144; Conservative 0; Mismatches 161; Indels 0; Gaps 0;
QY 218 TTCCTCAAGGTAAGAGCTGAGTATAAAGCTTCAACTGATTTTAAATTTCTTTTACGA 277
Db 611 TCCTTCATGTTAGAAACCTGGCTTTTACTACTACTACTACTACTACTACTACTACTA 670
QY 278 CAACTACTAATGAGGGAGAACATATGTAATAAAAAAGACTGCGTCAGCAAAATGAGA 337
Db 671 CTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 730
QY 338 TTGGACATGGGCTAAATCTATATCAGCTAATACTACACAGTTTCCACTGTACTGAGT 397
Db 731 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 790
QY 398 CAAATAATGATGCTACTGAGGTTTAAATGTTTCCCAATATGATATATTATGTTTCTA 457
Db 791 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 850
QY 458 GCACTGTTAATAATGAGGCTGTAATTTATGTTTACATCTGTAACCTCCAAATGCTACTATT 517
Db 851 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 910
QY 518 ATCAA 522
Db 911 CTGCA 915

RESULT 13
US-09-181-585-1
; Sequence 1, Application US/09181585
; Patent No. 6524791
; GENERAL INFORMATION:
; APPLICANT: Ranum, Laura P.W.
; APPLICANT: Koob, Michael
; TITLE OF INVENTION: SPINOCEREBELLAR ATAXIA TYPE 8 AND METHODS OF DETECTION
; FILE REFERENCE: 1100900101
; CURRENT APPLICATION NUMBER: US/09/181,585
; CURRENT FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1159
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-181-585-1

Query Match 3.1%; Score 47.4; DB 4; Length 1159;
Best Local Similarity 47.2%; Pred. No. 0.016;
Matches 144; Conservative 0; Mismatches 161; Indels 0; Gaps 0;
QY 218 TTCCTCAAGGTAAGAGCTGAGTATAAAGCTTCAACTGATTTTAAATTTCTTTTACGA 277
Db 423 TCCTTCATGTTAGAAACCTGGCTTTTACTACTACTACTACTACTACTACTACTACTA 482
QY 278 CAACTACTAATGAGGGAGAACATATGTAATAAAAAAGACTGCGTCAGCAAAATGAGA 337

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

QM nucleic - nucleic search, using sw model

Run on: September 13, 2004, 01:51:02 ; Search time 730.222 Seconds
(without alignments)
10397.614 Million cell updates/sec

Title: us-10-009-254-1

Perfect score: 1509

Sequence: 1 atgaaaagaaatgattca.....gtcgctgattacgtttcttaa 1509

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3304383 seqs, 2515761380 residues

Total number of hits satisfying chosen parameters: 6608766

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*

2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*

3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*

4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*

5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*

6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*

7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*

8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*

9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*

10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*

11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*

12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*

13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*

14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*

15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*

16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*

17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*

18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*

19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match %	Score	Length	ID	Description
1	1509	100.0	1509	12	US-10-333-002-15
2	64.8	4.3	1983	13	US-10-206-576-217
3	64.8	4.3	2199	12	US-10-333-002-28
4	64.8	4.3	15614	9	US-09-070-927A-45
5	63.2	4.2	1881	17	US-10-661-809-12
6	59.2	3.9	1137	13	US-10-282-122A-15988
7	53	3.5	4985	15	US-10-094-240-10
8	53	3.5	4985	15	US-10-056-405-10
9	50.4	3.3	14066	15	US-10-349-680-149
10	50.4	3.3	14067	13	US-10-282-122A-40681
11	50	3.3	3996	15	US-10-087-464-42
12	49.8	3.3	1687	13	US-10-206-576-219
13	48.6	3.2	4997	13	US-10-282-122A-35506
14	48.4	3.2	3666	9	US-09-137-531-13

15	48.4	3.2	3666	9	US-09-137-531-14	Sequence 14, Appl
16	48.4	3.2	4197	9	US-09-137-531-7	Sequence 7, Appl
17	48.4	3.2	4197	9	US-09-137-531-8	Sequence 8, Appl
18	47.4	3.1	1037	16	US-10-373-667-3	Sequence 3, Appl
19	47.4	3.1	1159	16	US-10-373-667-1	Sequence 1, Appl
20	47.4	3.1	1471	16	US-10-373-667-2	Sequence 2, Appl
21	47.4	3.1	4047	9	US-09-815-242-4843	Sequence 4843, Ap
22	47.4	3.1	4050	9	US-09-815-242-9039	Sequence 9039, Ap
23	47.4	3.1	4358	8	US-08-781-986A-454	Sequence 454, App
24	47.4	3.1	4358	13	US-10-329-624-454	Sequence 454, App
25	47.2	3.1	5314	15	US-10-155-533-1	Sequence 1, Appl
26	46.8	3.1	2801	13	US-10-282-122A-34844	Sequence 34844, A
27	46.8	3.1	7047	15	US-10-240-453-260	Sequence 260, App
28	46.2	3.1	4158	13	US-10-282-122A-8140	Sequence 8140, Ap
29	45.8	3.0	684707	16	US-10-398-221-9	Sequence 9, Appl
30	45.8	3.0	3011208	16	US-10-398-221-2058	Sequence 2058, Ap
31	45.6	3.0	2017	15	US-10-155-533-3	Sequence 3, Appl
32	45.6	3.0	6397	13	US-10-221-714A-181	Sequence 181, App
33	45.6	3.0	6397	15	US-10-239-676-107	Sequence 107, App
34	45.6	3.0	6397	15	US-10-311-455-1319	Sequence 1319, Ap
35	45.6	3.0	6397	15	US-10-240-453-119	Sequence 119, App
36	45.4	3.0	2217	13	US-10-282-122A-17439	Sequence 17439, A
37	45.2	3.0	3931	15	US-10-006-780-1	Sequence 1, Appl
38	45	3.0	12237	15	US-10-311-455-2331	Sequence 2331, Ap
39	44.8	3.0	573	15	US-10-029-386-25433	Sequence 25433, A
40	44.8	3.0	5358	17	US-10-637-544-19	Sequence 19, Appl
41	44.8	3.0	16287	15	US-10-311-455-645	Sequence 645, App
42	44.6	3.0	954	17	US-10-451-467A-537	Sequence 537, App
43	44.4	2.9	7446	13	US-10-282-122A-7578	Sequence 7578, Ap
44	44.2	2.9	486	13	US-10-465-217-13	Sequence 13, Appl
45	44.2	2.9	606	15	US-10-349-680-85	Sequence 85, Appl

ALIGNMENTS

RESULT 1

US-10-333-002-15
; Sequence 15, Application US/10333002
; Publication No. US20040071729A1

; GENERAL INFORMATION:

; APPLICANT: Adderson, Elisabeth

; APPLICANT: Bohnsack, John

; TITLE OF INVENTION: Group B Streptococcus Polypeptides Nucleic Acids and Therapeutic
; TITLE OF INVENTION: Compositions and Vaccines Thereof

; FILE REFERENCE: 2511-1-001 (SJ-0039)

; CURRENT APPLICATION NUMBER: US/10/333,002

; CURRENT FILING DATE: 2003-07-08

; PRIOR APPLICATION NUMBER: PCT/US01/24795

; PRIOR FILING DATE: 2001-08-08

; PRIOR APPLICATION NUMBER: US 09/634,341

; PRIOR FILING DATE: 2000-08-08

; NUMBER OF SEQ ID NOS: 37

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 15

; LENGTH: 1509

; TYPE: DNA

; ORGANISM: Streptococcus agalactiae

US-10-333-002-15

Query Match 100.0%; Score 1509; DB 12; Length 1509;
Best Local Similarity 100.0%; Pred. No. 6.8e-314;
Matches 1509; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATCAAAAGAAATGATTCAATCGCTGTAGTCGCGAGTTAGCATTTGGTATGCGCTGTA 60

Db 1 ATCAAAAGAAATGATTCAATCGCTGTAGTCGCGAGTTAGCATTTGGTATGCGCTGTA 60

Qy 61 TCACCAAGTACGCGCGATAGCTTTTTCGCGCTGAGACAGGGAACAATACAGTTCAAGTACT 120

Db 61 TCACCAAGTACGCGCGATAGCTTTTTCGCGCTGAGACAGGGAACAATACAGTTCAAGTACT 120

Qy 121 CAAAAAGGCGCAACCTATAAAGCATATAAAGTTTTTGTATGACAGAAATAGATATGCAAT 180

Db 121 CAAAAGGCGCAACCTATAAAGCATATAAAGTTTGTGACAGAAATAGATAATGCAAT 180
Qy 181 GTATCTGATTCGAATAAAGAGTGGAGCTCTCTATTAATTCCTCAAGGTAAAGAGCTGAG 240
Db 181 GTATCTGATTCGAATAAAGAGTGGAGCTCTCTATTAATTCCTCAAGGTAAAGAGCTGAG 240
Qy 241 TATAAGGCTTCAACTGATTTTAAATCTCTCTTTTACGACAACTACTAATGGAGGAGAAC 300
Db 241 TATAAGGCTTCAACTGATTTTAAATCTCTCTTTTACGACAACTACTAATGGAGGAGAAC 300
Qy 301 TATGTAACTAAAAGATACCTGGTCAGCAAAATGAGATTGCGCATGGCTTAATCTATA 360
Db 301 TATGTAACTAAAAGATACCTGGTCAGCAAAATGAGATTGCGCATGGCTTAATCTATA 360
Qy 361 TCAGCTAATACTACACCACTTTTCACTGTTACTGAGTCAAAATATGATGGTACTGAGGTT 420
Db 361 TCAGCTAATACTACACCACTTTTCACTGTTACTGAGTCAAAATATGATGGTACTGAGGTT 420
Qy 421 ATTAATGTTTCCCAATATGGATATTAATATGTTTCTAGCACTGTTAATATGGAGCTGTA 480
Db 421 ATTAATGTTTCCCAATATGGATATTAATATGTTTCTAGCACTGTTAATATGGAGCTGTA 480
Qy 481 ATTATGTTTACATCTGTAACCTCCAAATGCTACTATTTCATGAAAGAAATACATGCGACA 540
Db 481 ATTATGTTTACATCTGTAACCTCCAAATGCTACTATTTCATGAAAGAAATACATGCGACA 540
Qy 541 TGGGAGATGTTGGTGGAAAACTGTAGTCAAAAAAGTACTCGTGGTGGTGTATACAGTC 600
Db 541 TGGGAGATGTTGGTGGAAAACTGTAGTCAAAAAAGTACTCGTGGTGGTGTATACAGTC 600
Qy 601 AAATATATCTATTCTATAGAATGCGATCAATATCATGGTACAGAAAAAGTGTATCAA 660
Db 601 AAATATATCTATTCTATAGAATGCGATCAATATCATGGTACAGAAAAAGTGTATCAA 660
Qy 661 TATGTTATAAGGATACATGCGATCTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTAT 720
Db 661 TATGTTATAAGGATACATGCGATCTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTAT 720
Qy 721 GAAGTAACTATTACTGATGATCAGGGAATATTACAACTCTAATCAAGGTTCGGAAAAA 780
Db 721 GAAGTAACTATTACTGATGATCAGGGAATATTACAACTCTAATCAAGGTTCGGAAAAA 780
Qy 781 GCACTGGGAAGTATACCTGTTAGAGGAAATTAATTTTACAGATTACTATTCGGTG 840
Db 781 GCACTGGGAAGTATACCTGTTAGAGGAAATTAATTTTACAGATTACTATTCGGTG 840
Qy 841 GCAGCTACCAATACTCCAAACCGGAATACTCAAAATCGAGCTAATGATGACTTTTTTAT 900
Db 841 GCAGCTACCAATACTCCAAACCGGAATACTCAAAATCGAGCTAATGATGACTTTTTTAT 900
Qy 901 AAGGGAATAAATCAATCAAGTCACTTTATACAGGAGTATTAAGAGTGGAGCTAAACCA 960
Db 901 AAGGGAATAAATCAATCAAGTCACTTTATACAGGAGTATTAAGAGTGGAGCTAAACCA 960
Qy 961 GGTTCAGTCAATTTACAGAAATACAAATTCGACCATCAATCAATCAATCAATCAAT 1020
Db 961 GGTTCAGTCAATTTACAGAAATACAAATTCGACCATCAATCAATCAATCAATCAAT 1020
Qy 1021 GATGACCCAGGTCAAAAGTAAACAGTGGAGGATGGTCAAAATTTACTATAAAAAAATGAT 1080
Db 1021 GATGACCCAGGTCAAAAGTAAACAGTGGAGGATGGTCAAAATTTACTATAAAAAAATGAT 1080
Qy 1081 GGTTCACAAAAGTTCATTAAGGTGCTATATTTGTTTAAAGAAATGCTACGGTCAA 1140
Db 1081 GGTTCACAAAAGTTCATTAAGGTGCTATATTTGTTTAAAGAAATGCTACGGTCAA 1140
Qy 1141 TTTCTAACTTTAAGGATCAAAATACGTTGATGGGACAGAGCTAATGCAACAGAA 1200
Db 1141 TTTCTAACTTTAAGGATCAAAATACGTTGATGGGACAGAGCTAATGCAACAGAA 1200
Qy 1201 TATACACAGGAGCAGATGGTATATTAATCAATACAGGCTTGAAGAGAGTACATACTAT 1260

Db 1201 TATACACAGGAGCAGATGGTATATTAATCAATACAGGCTTGAAGAGGATACATACTAT 1260
Qy 1261 CTAGTTGAGAAAAAGGCTCCCTTAGGTACAAATTTGTTAGTAACTCTCAGAGGTTATT 1320
Db 1261 CTAGTTGAGAAAAAGGCTCCCTTAGGTACAAATTTGTTAGTAACTCTCAGAGGTTATT 1320
Qy 1321 TTAGGAGATCGAGCCACTGATACGACTAAATTCAGATAACCTTTTAGTTAACCCAACTGTT 1380
Db 1321 TTAGGAGATCGAGCCACTGATACGACTAAATTCAGATAACCTTTTAGTTAACCCAACTGTT 1380
Qy 1381 GAAAATAACAAAGTACTGAGTTGCCCTTCAACAGGTTGTTGTTGTTGTTGTTGTTGTTA 1440
Db 1381 GAAAATAACAAAGTACTGAGTTGCCCTTCAACAGGTTGTTGTTGTTGTTGTTGTTGTTA 1440
Qy 1441 ATTATAGTGCATTTTACTGATTAAGGAGCAGGATCGTGGCTTTGTTGTTGTTGTTGTTA 1500
Db 1441 ATTATAGTGCATTTTACTGATTAAGGAGCAGGATCGTGGCTTTGTTGTTGTTGTTGTTA 1500
Qy 1501 CGTTCTTAA 1509
Db 1501 CGTTCTTAA 1509

RESULT 2
US-10-206-576-217
; Sequence 217, Application US/10206576
; Publication No. US20030017495A1
; GENERAL INFORMATION:
; APPLICANT: Choi et al.
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 497
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-R
; COMPUTER: Dell Latitude
; OPERATING SYSTEM: Windows 98
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/206,576
; FILING DATE: 29-Jul-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/071,035
; FILING DATE: 1998-05-04
; APPLICATION NUMBER: US 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: US 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: US 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Hyman, Mark J.
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB369PID1
; INFORMATION FOR SEQ ID NO: 217:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1983 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 217:
US-10-206-576-217

Query Match 4.3%; Score 64.8; DB 13; Length 1983;
Best Local Similarity 52.1%; Pred. No. 0.0013;
Matches 172; Conservative 0; Mismatches 152; Indels 6; Gaps 1;

QY 1180 ACAGAGCTAATCAACAGATATACACAGGAGCAGATGGTATATAATTACCATTCAGGC 1239
Db 1660 ACAAAGCTGAAGCAACTCTTTTACAAACCGCTGATGGATTAGTGTATATACAGGG 1719
QY 1240 TTGAAAGAGGTACATACTATCTAGTTGAGAAAAAGGCTCCCTTAGGTACAATTGTTA 1299
Db 1720 CTTAATACGGTACCTATTATTATTAGAGAACTGTAGTCTCTGATGATTGCTTGTTA 1779
QY 1300 GATACTCTCAGAAGGTATTTTAGGAGATGGAGCCACTGATACGACTAATTCAGATAAC 1359
Db 1780 ACAATCGGATTGAATTTGGTCAATGAACAATCATATGCGCAA-----CAGAAAC 1833
QY 1360 CTTTGTAGTAAACCACTGTGAAATAACAAAGGTACTGAGTTGCTTCAACAGGTGGT 1419
Db 1834 CTAGTTTACCAGAAAAGTACCACAAACAAACAAAGGTACTTACCTTCAACAGGTGGC 1893
QY 1420 ATTGGTACAAATTTTCTACATTATAGGTGCAATTTTAGTAATAGGAGCAGGTATCGTG 1479
Db 1894 AAAGGAATCTACGTTTACTTAGGAGTGGCGAGTCTTGCTACTATTGAGGAGTCTAC 1953
QY 1480 CTTGTTGCTCGTCTGCTGTTTACGTTCTTAA 1509
Db 1954 TTTGCTAGACGTAGAAAAGAAATGCTTAA 1983

RESULT 3

US-10-333-002-28
; Sequence 28, Application US/10333002
; Publication No. US20040071729A1
; GENERAL INFORMATION:
; APPLICANT: Adderson, Elisabeth
; TITLE OF INVENTION: Group B Streptococcus Polypeptides Nucleic Acids and Therapeutic
; FILE REFERENCE: 2511-1-001 (SJ-0039)
; CURRENT APPLICATION NUMBER: US/10/333,002
; CURRENT FILING DATE: 2003-07-08
; PRIOR APPLICATION NUMBER: PCT/US01/24795
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 09/634,341
; PRIOR FILING DATE: 2000-08-08
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-10-333-002-28

Query Match 4.3%; Score 64.8; DB 12; Length 2199;
Best Local Similarity 52.1%; Pred. No. 0.0013;
Matches 172; Conservative 0; Mismatches 152; Indels 6; Gaps 1;
QY 1180 ACAGAGCTAATCAACAGATATACACAGGAGCAGATGGTATATAATTACCATTCAGGC 1239
Db 275 ACAAAGCTGAAGCAACTCTTTTACAAACCGCTGATGGATTAGTGTATATACAGGG 334
QY 1240 TTGAAAGAGGTACATACTATCTAGTTGAGAAAAAGGCTCCCTTAGGTACAATTGTTA 1299
Db 335 CTTAATACGGTACCTATTATTATTAGAGAACTGTAGTCTCTGATGATTGCTTGTTA 394
QY 1300 GATACTCTCAGAAGGTATTTTAGGAGATGGAGCCACTGATACGACTAATTCAGATAAC 1359
Db 395 ACAATCGGATTGAATTTGGTCAATGAACAATCATATGCGCAA-----CAGAAAC 448
QY 1360 CTTTGTAGTAAACCACTGTGAAATAACAAAGGTACTGAGTTGCTTCAACAGGTGGT 1419
Db 449 CTAGTTTACCAGAAAAGTACCACAAACAAACAAAGGTACTTACCTTCAACAGGTGGC 508
QY 1420 ATTGGTACAAATTTTCTACATTATAGGTGCAATTTTAGTAATAGGAGCAGGTATCGTG 1479
Db 509 AAAGGAATCTACGTTTACTTAGGAGTGGCGAGTCTTGCTACTATTGAGGAGTCTAC 568

QY 1480 CTTGTTGCTCGTCTGCTGTTTACGTTCTTAA 1509
Db 569 TTTGCTAGACGTAGAAAAGAAATGCTTAA 598

RESULT 4

US-09-070-927A-45
; Sequence 45, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
; Patrick J. Dillon
; Steven Barash
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,927A
; FILING DATE: 04-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 45:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15614 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 45:
US-09-070-927A-45

Query Match 4.3%; Score 64.8; DB 9; Length 15614;
Best Local Similarity 52.1%; Pred. No. 0.0035;
Matches 172; Conservative 0; Mismatches 152; Indels 6; Gaps 1;
QY 1180 ACAGAGCTAATCAACAGATATACACAGGAGCAGATGGTATATAATTACCATTCAGGC 1239
Db 13724 ACAAAGCTGAAGCAACTCTTTTACAAACCGCTGATGGATTAGTGTATATACAGGG 13783
QY 1240 TTGAAAGAGGTACATACTATCTAGTTGAGAAAAAGGCTCCCTTAGGTACAATTGTTA 1299
Db 13784 CTTAATACGGTACCTATTATTATTAGAAAGAACTGTAGTCTCTGATGATTGCTTGTTA 13843
QY 1300 GATACTCTCAGAGGTATTTTAGGAGATGGAGCCACTGATACGACTAATTCAGATAAC 1359
Db 13844 ACAATCGGATTGAATTTGGTCAATGAACAATCATATGCGCAA-----CAGAAAC 13897
QY 1360 CTTTGTAGTAAACCACTGTGAAATAACAAAGGTACTGAGTTGCTTCAACAGGTGGT 1419
Db 13898 CTAGTTTACCAGAAAAGTACCACAAACAAACAAAGGTACTTACCTTCAACAGGTGGC 13957

Qy 1420 ATTGGTACAACAATTTTCTACATTAATAGGTGCAATTTTATAGGACGAGTATCGTG 1479

Db 13958 AAGGAATCTACGTTTACTTTAGGAAGTGGCGAGCTTGTCTACTTATTGCAGAGTCTAC 14017

Qy 1480 CTGTGCTCGTCTGCTGTTTACGTTCTTAA 1509

Db 14018 TTGTGTAGACGTAGAAAAGAAATGCTTAA 14047

RESULT 5

```

US-10-661-809-12
; Sequence 12, Application US/10661809
; Publication No. US200401019A1
; GENERAL INFORMATION:
; APPLICANT: HOOK, Magnus
; TITLE OF INVENTION: BIOINFORMATIC METHOD FOR IDENTIFYING SURFACE-ANCHORED PROTEINS
; FILE REFERENCE: F07741US01/BAS
; CURRENT APPLICATION NUMBER: US/10/661,809
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: 60/410303
; PRIOR FILING DATE: 2002-09-13
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 1881
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-10-661-809-12

```

Query Match 4.2%; Score 63.2; DB 17; Length 1881;
Best Local Similarity 52.5%; Pred. No. 0.0027;
Matches 165; Conservative 0; Mismatches 143; Indels 6; Gaps 1;

1180	QY	ACAGAGCTAAGCAACAGAAATATACACAGGACAGATGGTATTAATTAACCATTCACAGC	1239
1561	Db	ACAAAGCTGAAGCAACTACTTTTCAACAAACGGCTGATGGATTAGTTGATATCACAGG	1620
1240	QY	TTGAAAGAGGTACATACTACTCTAGTTTGAGAAAAGCGTCCCTTAGGTTCACAAATTCGTTA	1299
1621	Db	CTTAATACGGTACTATTATTAGAGAAACCTGTAGCTCCCTGATGTTATGCTCTTGTTA	1680
1300	QY	GATTAAGCTCAGAAGGTTATTTTAGGAGATGGAGCCACTGATACGACTAAATTCAGATAAC	1359
1681	Db	ACAAATCGGATGGAATTTGGTCAATAGAACAAATCATATGGCACAA-----CAGAAAAC	1734
1360	QY	CTTTTAGTTAAACCAACTGTTTGAATAATACAAAGGTACTAGCTGCCTTTCACAGGTGGT	1419
1735	Db	CTAGTTTCACGAGAAAAGTACCACACAAACACAAAGGTACCTTACCTTCAACAGGTGGC	1794
1420	QY	ATTGGTACACAAATTTCTACATTATAGGTGCATTTTAGTAATAGGACGAGGTATCGTG	1479
1795	Db	AAAGGAATCTACGTTTACTTAGGAAGTGGCGAGTCTTGCTACTTATTTCAGGAGTCTAC	1854
1480	QY	CTTGTTGCTCGTCG	1493
1855	Db	TTTGCTAGACGTAG	1868

RESULT 6

```

US/10-282-132A-15988
; Sequence 15988, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlssen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John

```

```

/ APPLICANT: Carr, Grant
/ APPLICANT: Yamamoto, Robert
/ APPLICANT: Forsyth, R.
/ APPLICANT: Xu, H.
/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
/ FILE REFERENCE: ELITFA.034A
/ CURRENT APPLICATION NUMBER: US/10/282,122A
/ CURRENT FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 15988
/ LENGTH: 1137
/ TYPE: DNA
/ ORGANISM: Clostridium acetobutylicum
US-10-282-122A-15988

```

Query Match 3.9%; Score 59.2; DB 13; Length 1137;
Best Local Similarity 43.7%; Pred. No. 0.015;
Matches 262; Conservative 0; Mismatches 338; Indels 0; Gaps 0;

Qy		90	TGAGACAGGGACAATTA	CAGTTCAAGATACTCAA	AAGCGCCAACTTATAA	GCATATAA	149
Db		30	TGGGACAGGCCAGAAGCT	ATAAAGATGGCTCCTTT	TGGTTAAAAAAC	TGAAGAAAATAA	89
Qy		150	AGTTTTTGATGCAGAAAT	TAGATAATGCAATGTAT	CTGATTGCAATAAAGAT	GAGCTTC	209
Db		90	AAATTTTCAGCAAAAGTAT	GTGTTACAGCACGCA	CAGGGAGATGCTAGAT	CAAGTTCT	149
Qy		210	TTATTTAATTCCTCAAGT	TAAGAAGCTGAGTATAA	AGCTTCCA	ACTGATTTTAATTC	269
Db		150	TAGTTTGTTCATATAGAG	CCAGACTTTTGATTAAT	TAAATGAAANAATG	CAATCTCT	209
Qy		270	TTTTTAGCACACTAAT	TGAGGGAGAACATATG	TAACTAAAAAAAGAT	CTCGGTCAGC	329
Db		210	TACAACGATAACTCCAGT	GTGTTTCA	TGGGTTAAGCGAAATATTT	TGCTGTAAGACG	269
Qy		330	AAATGAGATTGCGACAT	GGGCTAAATCTATAT	CAGCTTAATAC	TACCACCTGTTCCACTCT	389
Db		270	AGATATTATTTAGTACAT	GGAGATACAACTAC	ACGTTTGCAGCATCACT	TGCAGCCTT	329
Qy		390	TACTGAGTCAATAAT	AGATGGTACTGAGGTAT	TAAATGTTTCCCAATAT	TGGATATPATTA	449
Db		330	TTATGAAAGATAGCTAT	TGGGCATGTG	TGAGGCTGGACCTTAGAA	CTTATGATAAATAATTT	389
Qy		450	TGTTCTTAGCACTGTT	TAATATGAGCTGTAAT	TATGTTTACATCTGTA	ACTCCAAATGC	509
Db		390	TCCTTTTCTGAGAGTAA	ATATAGAAAACCTTACTG	GAGCAATG	GCAGATATGCAATTTTCG	449
Qy		510	TACTATTCAATGAAAG	AATACTGATGCGACAT	GGGAGATGGTGGT	TGGAAAAA	569
Db		450	ACCGACAGTAAATCT	TAATAATCTTTTAA	GAGAAAGGTGTAA	AGAAAAAATAATTT	509

QY 570 TCAAAAACCTACTCGGTTGGTATACAGTCAATATATCTATTACTTATTAAGAATCGAGT 629
DB 510 TATAACTGGAAATACGTGATGCTATGAAATACACAGTGGATAGTAATTTATGTAAT 569
QY 630 CAATTATCATGGTACAGAAAAGTGTATCAATATGTTTATAAGGATACTATGCCATCTGC 689
DB 570 TAAAATGATCAATTAATAAATGGACTACAAACATATAAAGATTTATTATGTTAACTGC 629

RESULT 7

US-10-094-240-10/c
; Sequence 10, Application US/10094240
; Publication No. US20030082637A1
; GENERAL INFORMATION:
; APPLICANT: ZWIEBEL, LAURENCE J.
; TITLE OF INVENTION: ARRESTIN GENE, POLYPEPTIDE, AND METHODS OF USE THEREOF
; FILE REFERENCE: N8289
; CURRENT APPLICATION NUMBER: US/10/094,240
; CURRENT FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 10/056,405
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: 60/264,649
; PRIOR FILING DATE: 2001-01-26
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 4985
; TYPE: DNA
; ORGANISM: Anopheles gambiae
US-10-094-240-10

Query Match 3.5%; Score 53; DB 15; Length 4985;
Best Local Similarity 44.9%; Pred. No. 0.68;
Matches 241; Conservative 0; Mismatches 295; Indels 1; Gaps 1;

QY 170 ATATGCAATGTATCTGATTCGAATAAAGATCGAGCTTCTTATTTAAATTCCTCAAGGTA 229
DB 2854 ATATAATATTAAATAATTAATTAATAATAATAATAATAATAATAATAATAATAATA 2795
QY 230 AAGAAGCTGATGATAAAGCTTCAACTGATTTTAAATTCCTTTTACGACAACTACTAATG 289
DB 2794 ATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 2735
QY 290 GAGGGAGACATATGTAACTAAATAAGATACTGGTCAGCAAAATGAGATTGCGACATGGG 349
DB 2734 TGATAATAATGTATAATAATGAATAACAATAATAATAATAATAATAATAATAATAA 2675
QY 350 CTAAATCTATATCA-GCTAATACTACACAGTTTCCAGTGTACTGAGTCABAATAATGAT 408
DB 2674 TTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 2615
QY 409 GGTACTGAGGTTATTAATGTTTCCCAATATGGATATTTATTTCTTCTAGCACTGTTAAT 468
DB 2614 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 2555
QY 469 AATGGAGCTGTAATATGTTTACATCTGTAACCTCCAAATGCTACTATTCATGAAAAGAA 528
DB 2554 AATAATAATAATAATAATAATAAAGTAATAATAATAATAATAATAATAATAATAATAA 2495
QY 529 ACTGATCGGACATGGGGAGATGGTGGGAAACTGTAGATCAAAAACGTACTCGGTT 588
DB 2494 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 2435
QY 589 GGTGATACAGTCAAAATATCTATTACTTATAAGAATGCAGTCAATTTATCATGTACAGAA 648
DB 2434 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 2375
QY 649 AAGTGTATCAATATGTTTATAAAGGATCTATGCCATCTGCTTCTGTAGTTGATTG 705
DB 2374 AATAATAATAATAATAATAATAAATCTGTATTTTGGTTCTCTGTAACATAATTG 2318

RESULT 8

US-10-349-680-149/c
; Sequence 149, Application US/10349680
; Publication No. US20030176654A1
; GENERAL INFORMATION:
; APPLICANT: Casell, Gail
; APPLICANT: Chen, Ellison
; APPLICANT: Glass, Jennifer
; APPLICANT: Glass, John
; APPLICANT: Heiner, Cheryl
; APPLICANT: Lefkowitz, Elliot
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREA PLASMA

US-10-056-405-10/c
; Sequence 10, Application US/10056405
; Publication No. US20030166013A1
; GENERAL INFORMATION:
; APPLICANT: ZWIEBEL, LAURENCE J.
; TITLE OF INVENTION: MOSQUITO OLFATORY GENES, POLYPEPTIDES, AND METHODS OF
; TITLE OF INVENTION: USE THEREOF
; FILE REFERENCE: N7841
; CURRENT APPLICATION NUMBER: US/10/056,405
; CURRENT FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: 60/264,649
; PRIOR FILING DATE: 2001-01-26
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 4985
; TYPE: DNA
; ORGANISM: Anopheles gambiae
US-10-056-405-10

Query Match 3.5%; Score 53; DB 15; Length 4985;
Best Local Similarity 44.9%; Pred. No. 0.68;
Matches 241; Conservative 0; Mismatches 295; Indels 1; Gaps 1;

QY 170 ATATGCAATGTATCTGATTCGAATAAAGATCGAGCTTCTTATTTAAATTCCTCAAGGTA 229
DB 2854 ATATAATATTAAATAATTAATTAATAATAATAATAATAATAATAATAATAATAATA 2795
QY 230 AAGAAGCTGATGATAAAGCTTCAACTGATTTTAAATTCCTTTTACGACAACTACTAATG 289
DB 2794 ATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 2735
QY 290 GAGGGAGACATATGTAACTAAATAAGATACTGGTCAGCAAAATGAGATTGCGACATGGG 349
DB 2734 TGATAATAATGTATAATAATGAATAACAATAATAATAATAATAATAATAATAATAA 2675
QY 350 CTAAATCTATATCA-GCTAATACTACACAGTTTCCAGTGTACTGAGTCABAATAATGAT 408
DB 2674 TTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 2615
QY 409 GGTACTGAGGTTATTAATGTTTCCCAATATGGATATTTATTTCTTCTAGCACTGTTAAT 468
DB 2614 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 2555
QY 469 AATGGAGCTGTAATATGTTTACATCTGTAACCTCCAAATGCTACTATTCATGAAAAGAA 528
DB 2554 AATAATAATAATAATAATAAAGTAATAATAATAATAATAATAATAATAATAATAATA 2495
QY 529 ACTGATCGGACATGGGGAGATGGTGGGAAACTGTAGATCAAAAACGTACTCGGTT 588
DB 2494 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 2435
QY 589 GGTGATACAGTCAAAATATCTATTACTTATAAGAATGCAGTCAATTTATCATGTACAGAA 648
DB 2434 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 2375
QY 649 AAGTGTATCAATATGTTTATAAAGGATCTATGCCATCTGCTTCTGTAGTTGATTG 705
DB 2374 AATAATAATAATAATAATAATAAATCTGTATTTTGGTTCTCTGTAACATAATTG 2318

RESULT 9

US-10-349-680-149/c
; Sequence 149, Application US/10349680
; Publication No. US20030176654A1
; GENERAL INFORMATION:
; APPLICANT: Casell, Gail
; APPLICANT: Chen, Ellison
; APPLICANT: Glass, Jennifer
; APPLICANT: Glass, John
; APPLICANT: Heiner, Cheryl
; APPLICANT: Lefkowitz, Elliot
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREA PLASMA

; TITLE OF INVENTION: UREALYTICUM
; FILE REFERENCE: UAB-13403/22
; CURRENT APPLICATION NUMBER: US/10/349,680
; CURRENT FILING DATE: 2003-01-23
; PRIOR APPLICATION NUMBER: US 09/601,198
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: PCT/US99/01972
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: US 60/073,189
; PRIOR FILING DATE: 1998-01-30
; NUMBER OF SEQ ID NOS: 181
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 149
; LENGTH: 14066
; TYPE: DNA
; ORGANISM: Ureaplasma urealyticum
US-10-349-680-149

Query Match 3.3%; Score 50.4; DB 15; Length 14066;
Best Local Similarity 42.8%; Pred. No. 4.1;
Matches 313; Conservative 0; Mismatches 416; Indels 3; Gaps 1;
QY 435 ATATCGATATTATTATTTCTAGCACGTGTAATAATGAGCTGTAATTATGTTTACATC 494
Db |||||
QY 5284 AAATGTTAGTTCTAATGATTTCTAATAATCTACACACAAACAGTTAATATTGATTTGA 5225
Db |||||
QY 495 TGTAACTCCAAATGCTACTATTTCATGAAAAGAACTGATCGACATGGGGAGATGGTG 554
Db |||||
QY 5224 TGGAAATCAACGAACCTGAATTAATAAAAAAATCAATTGTTATATCTTCAATGACGG 5165
Db |||||
QY 555 TGGAAAACCTGATGATCAAAAAAGCTACTCGTTGGTGATACAGTCAAAATATCTATTAC 614
Db |||||
QY 5164 TGAAGAAATTTAAGTGATCAAAAAACATTCCTTTGGCTAATAACCACTATAGTTTCGA 5105
Db |||||
QY 615 TTATAAGAAATGCAGTCAATTAATCATGTTACAGAAAAAGTGTATCAATATGTTATAAGGA 674
Db |||||
QY 5104 ATTGCTAATTTAAACATATCTGTAATAATACATAAAGAAAGTAAATTAATTAAGA 5045
Db |||||
QY 675 TACTATGCCATCTGCTTCTGTAGTTGATTTGAAAGGAGTCTTATGAAGTAACATATTAC 734
Db |||||
QY 5044 TGATAATAAAACATCAATAAATTTCCATTTAAAAAATGGAATAGCTGATTTATTGT 4985
Db |||||
QY 735 TGATGGATCAGGGAATATTAACATCTAACTCAAGTTTCGAAAAACCACTGGGAAGTA 794
Db |||||
QY 4984 TAATAAACAAATCGATTTAGTATAAGTCAATTAATGAGCCAAATTCAGAGCGAAAAA 4925
Db |||||
QY 795 TAACCTGTTAGAGAAATAATAATTTTCAAGTTACTATTCCGTTGGGCGAGTCAACATAC 854
Db |||||
QY 4924 TAATTACAAATCAACAAATTAGATTTATATTAATGATCCTGATATGTTTAAGTAA 4865
Db |||||
QY 855 TCCAAACCGGAATACTCAAAATGAGCTAATGATGACTTTTTTATAAGGGAATAAATAC 914
Db |||||
QY 4864 TGAAGAAGAACCAATCAATTAATTAATGTTGAGAAAT---TTATCTGTAAGCAAAAGTTAA 4808
Db |||||
QY 915 AATCACAGTCACTTATACAGGAGTATTAAGAGTGGAGCTAAACCCAGTTTCACTGATTT 974
Db |||||
QY 4807 AATTGTAGGAGTCAAAATATTTAGAAACAACATTCATTAATCTAGCATTAACCAAGA 4748
Db |||||
QY 975 ACCAGAAATAACAACTTGGACCAATCAACCCCAATPACTAGCAATGATGACCCAGGTCA 1034
Db |||||
QY 4747 TACTATTATTAACAGTATTACTTTTATTAATAACCAATAAAGCTGTACGAATATTGG 4688
Db |||||
QY 1035 ABAAGTAACTAGTGGGATGCTCAATTAATTAATAAATAAATAAATTTGATGGTTTCCACAAAGC 1094
Db |||||
QY 4687 AATAGATAATAGCAATAAATAATATAACAATAACAATAATAAATCCATTAAATAATTGA 4628
Db |||||
QY 1095 TTCAATCAAGGTTGCTATATTTTGTAAAGAAATGCTACGGGTCAATTTCTTAACCTTAA 1154
Db |||||
QY 4627 TAATAATTTTGTGTCATTGGTCTGTTCAAAGTGATCTAATAAAGCTATTACTTTGGA 4568
Db |||||
QY 1155 CGATCAATAATA 1166
Db |||||
QY 4567 ATATAATAATAA 4556

RESULT 10

US-10-282-122A-40681
; Sequence 40681, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITPA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 40681
; LENGTH: 14067
; TYPE: DNA
; ORGANISM: Ureaplasma urealyticum
US-10-282-122A-40681

Query Match 3.3%; Score 50.4; DB 13; Length 14067;
Best Local Similarity 42.8%; Pred. No. 4.1;
Matches 313; Conservative 0; Mismatches 416; Indels 3; Gaps 1;
QY 435 ATATCGATATTATTATTTCTAGCACCTGTTAATAATGAGCTGTAATTATGTTTACATC 494
Db |||||
QY 495 TGTAACTCCAAATGCTACTATTTCATGAAAAGAAATCTGATGCGACATGGGGAGATGGTG 554
Db |||||
QY 8844 TCGAAATTCACGAACCTTGAATTAATAAAAAAATTCAAATTTGGTATATCTTCAATGACGG 8903
QY 555 TCGAAAACCTGATAGATCAAAAACCTACTCGGTTCTGTATACAGTCAAAATATCTATTAC 614
Db |||||
QY 8904 TGAAGAAATTTTAACTGATCAAAAACATTCGTTGGCTAATAACCATATAGTTTGA 8963
QY 615 TTATAAGAAATCGAGTCAATTAATTCATGGTACAAAAAGTGTATCAATATGTTTATAAGGA 674
Db |||||
QY 8964 ATTGTCTAATTTAAAAACATAATCGTAATAATATACATAAAGAAAGTAAGAAATTTTAAATGA 9023
QY 675 TACTATGCCATCTGCTTCTGTAGTTGATTTGAAACGAAGGCTCTTATGAAGTAACTATTAC 734

Db 9024 TGATGAAATCGTAGTAAATTTTCCATTTTAAATAATGGAATAGCTGATTTATTTGT 9083
Qy 735 TGATGATCAGGGAATATTAACAACCTCTAATCAAGGTTTCGGAATAAGCACTCGGAAGTA 794
Db 9084 TAATAAACAACAATCGATTAGTATAAGTTCAATAAATGAGCCCAATTCGAAGAGCGAAAA 9143
Qy 795 TAACCTGTTAGAGGAATAATTAATTTTCAGATTAATTTCCGTTGGCAGCTACCAATAC 854
Db 9144 TAATTTTACAATCAACACAAATTTAGATTTATATTAATGATCCTGATATGTTTAAAGTAA 9203
Qy 855 TCCAAACCGGAAATACTCAAAATGGAGCTAATGATGACTTTTTTTTATAAGGGAATAATAC 914
Db 9204 TGAGAAGAAGCAATCAATTAATTTATGTTGAGAT---TTATCTGTAAAGCAAAAGTTAA 9260
Qy 915 AATCAGAGTCACCTTATACAGGAGTATTAAAGAGTGGAGCTAAACAGGTTTCAGCTGATTT 974
Db 9261 AATTGTAGGAGTCAGAAATATTATTAGAAACAACATTCATTAATCTAGCATTTAAACCAAGA 9320
Qy 975 ACCAGAAATACAAACATTCGGACCATCAACCCCAATAGCAATGATACCCAGGTCA 1034
Db 9321 TACTATTATTAACAGTATTAATTTTATTAATAAACCAATAAAGCTGCTACGAATATGG 9380
Qy 1035 AAAAGTAAACAGTGAAGGATGTCATTAATTTTACTATAAAAAAATGATGTTTCCACAAAGC 1094
Db 9381 AATAGATTAATAGCAATTAATAATATACAAATACAAACATTAATTAATCCATTAATAATGA 9440
Qy 1095 TTCATTACAAGTGCTATATTTGTTTAAAGAATGCTACGGGTCAATTTCTAAACCTTAA 1154
Db 9441 TAATAATTTTGTGTCATTGCTGCTGTTCAAAAGTGATCTAATAAAGTATTACTTTGGA 9500
Qy 1155 CGATACAATAA 1166
Db 9501 ATATAATAATAA 9512

RESULT 11

US-10-087-464-42
; Sequence 42, Application US/10087464
; Publication No. US20030059436A1
; GENERAL INFORMATION:
; APPLICANT: Chishti, Athar
; APPLICANT: Oh, Steven
; APPLICANT: Liu, David
; APPLICANT: Goel, Vikas
; APPLICANT: Li, Xuerong
; TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses Thereof
; FILE REFERENCE: S1237/7019
; CURRENT APPLICATION NUMBER: US/10/087,464
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 06/272,930
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 42
; LENGTH: 3996
; TYPE: DNA
; ORGANISM: Plasmodium falciparum
US-10-087-464-42

Query Match 3.3%; Score 50; DB 15; Length 3996;
Best Local Similarity 47.5%; Pred. No. 2.7;
Matches 149; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

Qy 366 TAATACTACACCAAGTTTCCACTGTTACTGAGTCAAAATGAATGGTACTGAGGTATTAA 425
Db 1881 TAATAATATGATGAGAATTTGTAATAATAATAATACCAATAATAATAATTAATAATA 1940
Qy 426 TGTTCCTCAATGATGATTAATTTATGTTTCTAGCAGCTGTTTAATAATGGAGCTGTAATAT 485
Db 1941 TAATAATAATTTATGTTAAACAATAATAATAACAATAATAATAAAGGACAATAATAA 2000
Qy 486 GGTTCATCTGTAACTCCAAATGCTACTATTTCATGAAAAGAAATCTGATCGGACATGGG 545

Db 2001 TGATGAAATCGTAGTAAATTTTCCATTTTAAATAATGGAATAGCTGATTTATTTGT 2060
Qy 546 AGATGTTGGTGGAAAACTGTAGATCAAAAAACGTACTCGTTGGTGTATACAGTCAAAATA 605
Db 2061 TGATGAAGATGATAACAATAATAATTAATGATGATGATAATATGAGTGAATAACGAAGAAAT 2120
Qy 606 TACTATTACTTATAAGNATGCAGTCAATTTATCATGTTACAGAAAAAGTGTATCAATATGT 665
Db 2121 GGAAGATAATGATGAAGATAACGATGAGTATAATAATAGTAATGATAGTTTATAAATATGA 2180
Qy 666 TATAAAGGATACTA 679
Db 2181 AGAAAAGATAGTA 2194

RESULT 12

US-10-206-576-219
; Sequence 219, Application US/10206576
; Publication No. US20030017495A1
; GENERAL INFORMATION:
; APPLICANT: Choi et al.
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 497
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-R
; COMPUTER: Dell Latitude
; OPERATING SYSTEM: Windows 98
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/206,576
; FILING DATE: 29-Jul-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/071,035
; FILING DATE: 1998-05-04
; APPLICATION NUMBER: US 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: US 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: US 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Hyman, Mark J.
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB369P1D1
; INFORMATION FOR SEQ ID NO: 219:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1687 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 219:
US-10-206-576-219

Query Match 3.3%; Score 49.8; DB 13; Length 1687;
Best Local Similarity 62.4%; Pred. No. 1.9;
Matches 78; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 1180 ACAGAGCTAATCCACAGAAATATACACAGGAGCAGTGTATTAATTAATACAGGC 1239
Db 1466 ACAAAGCTGAACCAACTCTTTTACAACAACCGCTGATGATAGTTGATATCACAGG 1525
Qy 1240 TTCAAGAGAGTACATCTACTATCTAGTTTGAGAAAAAGCTCCCTTAGGTTACAATTTCTTA 1299
Db 1526 CTTAATATCGGTACCTATTATTATTAGAGAAACCTGTAGCTCTCTGATGATGATTTGTTT 1585

QY 1300 GATAA 1304
Db 1586 ACAA 1590

RESULT 13
US-10-282-122A-35506
; Sequence 35506, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282.122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 35506
; LENGTH: 4997
; TYPE: DNA
; ORGANISM: Staphylococcus haemolyticus
US-10-282-122A-35506

Query Match 3.2%; Score 48.6; DB 13; Length 4997;
Best Local Similarity 43.5%; Pred. No. 6;
Matches 271; Conservative 0; Mismatches 349; Indels 3; Gaps 1;

QY 28 TTAGTGGCGAGTTAGCATTTGGTATGCTGCTATCACCAGTTACCGCGATAGCTTTGGC 87
Db 1741 TCAGATCAAAACNAGCAATTTGATACAGCAGTCTTAACCAAAATCATTAGTCAATAAT 1800

QY 88 GCTGAGACGGGCAATTTACAGTTTCAAGATACCTCAAAAAGCGCAACTATAAGCATAT 147
Db 1801 AACAAATGAGCCACTCAAGAGAGAGAGATGTAGCTTTAGCCAAATTTGATGAGCAGCA 1860

QY 148 AAGTTTTTGTGATGAGAAATAGATAATGCAAAATGATCTGATTCGAATAAAGATGAGCT 207
Db 1861 AAACAAGCTAAAGCTGCATAGATGCTGCAACTACAAATAATGC---TGATAGTGAAGCA 1917

QY 208 TCTTATTTAATCTCTCAAGGTAAGAGCTGAGTATATAAGCTTCACTGATTTTAATCT 267

Db 1918 ACAAACAATAACTACTACAATTTATTTCTGGAATACTTCTCTGATACGGTGAAGAAAGCAGCA 1977
QY 268 CTTTTTACGCAACTACTAATGAGGAGAGAACATATGTAACTAAAGATACTGCGTCA 327
Db 1978 GCAAGAAAAGCGATTGATGATGCGAGCAACTGCTAAAAAAGAGCGATAAATAATACGTCA 2037
QY 328 GCAAAATGAGATTGCGACATGGCTAAATCTATATCAGCTAATCTACTACACAGTTTCCACT 387
Db 2038 GATGCAACACAGAAGAAAAGATGAGCAATGCGAAAGTTGATGCGCTGTAACAGCT 2097
QY 388 GTTACTGAGTCAAAATATGATGCTAGTGGTATTATTAATGTTTCCCAATATGGATATTAT 447
Db 2098 GCAAAACAGCAATTTACACAAGCAACAAACAATGATAATGTAGACCAAGAACAAATAGC 2157
QY 448 TATGTTTCTAGCACTGTTAATAATGAGCTGTAATTAATGTTTCCCAATATGGATATTAT 507
Db 2158 GGTACTTCGACTATTACTGCTATACCAAGAGTTTACTAAAAAAGCAGCAGCAAGAAA 2217
QY 508 GCTACTATTTCATGAAAAGATACTGATGCCACATGGGAGATGCTGGTGAAAAAATCTGTA 567
Db 2218 GCAATTGATGATGAGTCGTTGCTAAGAAAGCAGCAATTGATCTGTTGCCGATGTACA 2277
QY 568 GATCAAAAACGTAAGTGGTGGTGTATACAGTCAAAATATATCTATTCTTATAAGATGCA 627
Db 2278 GATGAAGAAAACAGGCGAGCTAAAGATAAAGTTGATGCTGAAGCTACCAAGCAAGAAA 2337
QY 628 GTCAATTATCATGTCACAGAAA 650
Db 2338 GCGATTGATCAAGCTACTACAAA 2360

RESULT 14
US-09-137-531-13
; Sequence 13, Application US/09137531
; Patent No. US20020048916A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/137,531
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/682,517
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3666 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-137-531-13

Query Match 3.2%; Score 48.4; DB 9; Length 3666;
Best Local Similarity 44.5%; Pred. No. 5.6;
Matches 193; Conservative 0; Mismatches 241; Indels 0; Gaps 0;

QY 302 ATGTAACTAAAGATGCTGCTGAGCAAAATGAGATGCGACATGGCTAAATCTATAT 361
Db 2768 ATGTAAAGTCAACAACTGTTGATCTGCAACTGTTTCATTAAAGATAGTGCAATAAT 2827

QY 362 CAGCTAAATACTACACAGTTTCCACTGTTTACTGAGTCAAAATATGATGCTACTGAGGTTA 421
Db 2828 CATTATCTTACATTAGTTTGAACCTGGTGCTTAATACAGGTGTTTGTCTACACTGTTTC 2887

```
QY 422 TTAATGTTTCCCAATATGGATATATATATATGTTTCTAGCACTGTTAATAATGGAGCTGTA 481
Db 2888 AAGCTGTGTACATTATCTTTAACTGCTGGTACATTAAACAGTTACTTATGCGAGATGCTA 2947
QY 482 TTATGGTTTACATCTGTAACCTCCAAATGCTACTATTTCATGAAAAGAAATACCTGATGCCACAT 541
Db 2948 ABAATGCTGAGGTGTTGCTGAAAATATTACTGCTAGCGTAAACATTAAGAAAACACTACTG 3007
QY 542 GGGGAGATGGTGGGAAAACCTGTAGATCAAAAACGTACTCGGTTGGTGATACAGTCA 601
Db 3008 GAGCAATTAATCTTCTGATACATTTACACAAAGGTGATTACCATCAGCAGCTTACAGCAGCTG 3067
QY 602 AATATATCTATTACTTATAAGAAATGCAAGTCAATTTATCATGTGTACAGAAAAGTGTATCAAT 661
Db 3068 AATATATCTTCTAAATCAATTTGCTGCAATTTATACATTTGCAACAGGTGAAGGATTCACCT 3127
QY 662 ATGTTTAAAGGATACATATGCCATCTGCTTCTGTAGTTGATTGAAACGAAAGGCTTTATG 721
Db 3128 TAAATATTGATAATGCTGGTGCTCAAGTAATTAACCTTAGCAGGTAAAAGGTGCACAAG 3187
QY 722 AAGTAATTAATCT 735
Db 3188 GTGTAGCTGATGCT 3201
```

RESULT 15

```
US-09-137-531-14
; Sequence 14, Application US/09137531
; Patent No. US20020048816A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/137,531
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/682,517
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3666 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3666
US-09-137-531-14
```

```
Query Match 3.2%; Score 48.4; DB 9; Length 3666;
Best Local Similarity 44.5%; Pred. No. 5.6;
Matches 193; Conservative 0; Mismatches 241; Indels 0; Gaps 0;

QY 302 ATGTAATAAAAAGATACCTGCGTCAGCAAAATGAGATTGGCAGATGGGCTAAATCTATAT 361
Db 2768 ATGTAAGTGCACAACACTGTTGATACCTGCAACTGTTTCATTAAAAGATAGTGCAAAATAAT 2827
QY 362 CAGCTAATATACACCAAGTTTCACCTGTTACTAGTCAAAATATATGTTAGTACGAGTTA 421
Db 2828 CATTATCTCTTACATTAGTTGAAACCTGGTCTAATACAGGTGATTATTGCTACAACTGTT 2887
QY 422 TTAATGTTTCCCAATATGGATATATATATGTTTCTAGCACTGTTAATAATGGAGCTGTA 481
Db 2888 AAGCTGTGTACATTATCTTTAACTGCTGGTACATTAACAGTTACTTATGCGAGATGCTA 2947
```

```
QY 482 TTATGGTTTACATCTGTAACTCCAAATGCTACTATTTCATGAAAAGAAATACCTGATGCCACAT 541
Db 2948 ABAATGCTGAGGTGTTGCTGAAAATATTACTGCTAGCGTAAACATTAAGAAAACACTACTG 3007
QY 542 GGGGAGATGGTGGTGA AAAA AACTGTAGATCAAAAACGTACTCGGTTGGTGATACAGTCA 601
Db 3008 GAGCAATTAATCTTCTGATACATTTACACAAAGGTGATTACCATCAGCAGCTTACAGCAGCTG 3067
QY 602 AATATATCTATTACTTATAAGAAATGCAAGTCAATTTATCATGTGTACAGAAAAGTGTATCAAT 661
Db 3068 AATATATCTTCTAAATCAATTTGCTGCAATTTATACATTTGCAACAGGTGAAGGATTCACCT 3127
QY 662 ATGTTTAAAGGATACATATGCCATCTGCTTCTGTAGTTGATTGAAACGAAAGGCTTTATG 721
Db 3128 TAAATATTGATAATGCTGGTGCTCAAGTAATTAACCTTAGCAGGTAAAAGGTGCACAAG 3187
QY 722 AAGTAATTAATCT 735
Db 3188 GTGTAGCTGATGCT 3201
```

Search completed: September 13, 2004, 08:40:15
Job time : 733.222 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2004, 22:26:26 ; Search time 604.354 Seconds
(without alignments)
10607.240 Million cell updates/sec

Title: US-10-009-254-1

Perfect score: 1509

Sequence: 1 atgaaaagaagaatgattca.....gtcgtcggttaacgtttcttaa 1509

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseqn_29Jan04:*

1: Geneseqn1980s:*

2: Geneseqn1990s:*

3: Geneseqn2000s:*

4: Geneseqn2001as:*

5: Geneseqn2001bs:*

6: Geneseqn2002as:*

7: Geneseqn2003as:*

8: Geneseqn2003bs:*

9: Geneseqn2003cs:*

10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1509	100.0	1509	5 AAD02390	Aad02390 Virulent
2	1509	100.0	1509	6 ABK11585	Abk11585 S. agalac
3	64.8	4.3	1983	2 AAX20108	Aax20108 Enterococ
4	64.8	4.3	1983	6 ABN98093	Abn98093 E faecali
5	64.8	4.3	1983	7 ACA88057	ACA88057 E. faecal
6	64.8	4.3	1983	7 ABX61663	Abx61663 Enterococ
7	64.8	4.3	2199	6 ABK11591	Abk11591 E. faecal
8	64.8	4.3	15614	2 AAX12982	Aax12982 Enterococ
9	64.8	4.3	15614	6 ABS98777	Abs98777 Enterococ
10	59.6	3.9	11922	3 AAA70187	Aaa70187 Plasmodi
11	59.2	3.9	1137	7 ACA28118	ACA28118 Prokaryot
12	56.6	3.8	969	6 ABQ39490	Abq39490 Oligonuc
13	56.6	3.8	969	6 ABQ39491	Abq39491 Oligonuc
14	55.4	3.7	2408	4 ABL28164	AbL28164 Drosophil
15	55.2	3.7	32392	6 ABL56203	AbL56203 AmEPV gen
16	54	3.6	408	4 ABL28165	AbL28165 Drosophil
17	53.6	3.6	2535	4 ABL25476	AbL25476 Drosophil
18	53	3.5	4985	6 ABQ75107	AbQ75107 Anopheles
19	53	3.5	4985	9 ACF79720	ACF79720 Mosquito
20	53	3.5	6033	3 AAA70152	Aaa70152 Plasmodi
21	51.2	3.4	2703	6 ABN67916	Abn67916 Streptoco
22	51.2	3.4	110000	6 ABN71527_13	Continuation (14 o
23	50.4	3.3	14066	2 AAX99556	Aax99556 Nucleic a

24	50.4	3.3	14067	7 ACA52811	ACA52811 Prokaryot
25	50.2	3.3	7143	3 AAA70250	Aaa70250 Plasmodi
26	50	3.3	3996	6 AAD47004	Aad47004 Plasmodi
27	49.8	3.3	1687	2 AAX20109	Aax20109 Enterococ
28	49.8	3.3	1687	6 ABN98094	Abn98094 E faecali
29	49.8	3.3	1687	7 ACA88058	ACA88058 E. faecal
30	49.8	3.3	1687	7 ABX61664	Abx61664 Enterococ
31	49.6	3.3	3837	3 AAA70211	Aaa70211 Plasmodi
32	48.6	3.2	4997	7 ACA47636	ACA47636 Prokaryot
33	48.6	3.2	48551	6 AAS20800	Aas20800 Clostridi
34	48.4	3.2	4197	2 AAQ99430	Aaq99430 B. sphaer
35	47.8	3.2	740	4 AAC85918	Aac85918 rCP41 CDN
36	47.8	3.2	1083	5 AAS76745	Aas76745 DNA encod
37	47.8	3.2	8244	5 AAS66529	Aas66529 DNA encod
38	47.6	3.2	11143	4 ABL12834	AbL12834 Drosophil
39	47.4	3.1	1037	3 AAA59242	Aaa59242 Exons E,
40	47.4	3.1	1472	3 AAA59241	Aaa59241 Exons D,
41	47.4	3.1	2428	7 ADA89818	Ada89818 Staphyloc
42	47.4	3.1	3945	2 AAX77593	Aax77593 S. aureus
43	47.4	3.1	4047	4 AAS52261	Aas52261 Staphyloc
44	47.4	3.1	4047	7 ACF73996	ACf73996 Staphyloc
45	47.4	3.1	4050	4 AAS55402	Aas55402 Staphyloc

ALIGNMENTS

RESULT 1

AAD02390

ID AAD02390 standard; DNA; 1509 BP.

XX

AC AAD02390;

XX

DT 24-APR-2001 (first entry)

XX

DE Virulent group B Streptococcus agalactiae spbl DNA.

XX

KW Type III virulent group B; spbl; cell wall bound protein; antibacterial;
KW immunisation; group B streptococci; GBS infection; vaccine; bacteraemia;
KW pneumonia; meningitis; endocarditis; osteoarticular infection; ds.

XX

OS Streptococcus agalactiae.

XX

FH Key Location/Qualifiers

FT CDS 1..1509

FT /*tag= a

FT /product= "S. agalactiae spbl protein"

FT sig_peptide 19..87

FT /*tag= b

FT mat_peptide 88..1506

FT /*tag= c

FT /product= "S. agalactiae mature spbl protein"

FT WO200078787-A1.

FT 28-DEC-2000.

XX 21-JUN-2000; 2000WO-US017082.

XX 21-JUN-1999; 99US-0140084P.

XX (UTAH) UNIV UTAH RES FOUND.

XX Adderson E, Bohnsack J;

XX WPI; 2001-102693/11.

XX P-PSDB; AAY72357.

XX Polynucleotide from spbl and 2 genes derived from virulent Group B
XX streptococci, polypeptide encoded by the polynucleotide useful as vaccine
XX for immunizing a mammal against the streptococcal infection.

PS Claim 9; Page 20-23; 34pp; English.

XX The present sequence is spbl gene from type III virulent group B
 CC Streptococcus agalactiae. The spbl protein has the characteristics of a
 CC cell wall bound protein and has antibacterial activity. The N-terminus
 CC of the spbl protein is a hydrophilic, basic stretch of 6 amino acids
 CC followed by a 23 amino acid hydrophobic, proline rich core, consistent
 CC with a signal peptide. The hydrophilic mature protein terminates in
 CC atypical LexTG domain that immediately precedes a hydro- phobic 20 amino
 CC acid core and a short, basic hydrophilic terminus. The spbl protein is
 CC used as a vaccine to immunise mammals against group B Streptococci (GBS)
 CC infection (e.g. bacteraemia, pneumonia, meningitis, endocarditis and
 CC osteoarticular infections). Determination of the gene products specific
 CC to type III-3 GBS is useful for diagnosing mammals infected or colonised
 CC by virulent GBS
 XX

Sequence 1509 BP; 534 A; 241 C; 298 G; 436 T; 0 U; 0 Other;

Query Match 100.0%; Score 1509; DB 5; Length 1509;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1509; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAAAAAGAAATGATTCGCTGTAGTGGCGAGTTTACGATTTGGTATGGCTGTA 60
 DB 1 ATGAAAAAGAAATGATTCGCTGTAGTGGCGAGTTTACGATTTGGTATGGCTGTA 60

QY 61 TCACCAAGTACGCCGATAGCTTTCCCGCTGAGACAGGGA CAATTACAGTTCAAGATACT 120
 DB 61 TCACCAAGTACGCCGATAGCTTTCCCGCTGAGACAGGGA CAATTACAGTTCAAGATACT 120

QY 121 CAAAAAGGCGCACTATAAAGCATATAAAGTTTTGTATGATGAGAAATAGATAAATGCAAAAT 180
 DB 121 CAAAAAGGCGCACTATAAAGCATATAAAGTTTTGTATGATGAGAAATAGATAAATGCAAAAT 180

QY 181 GTATCTGATTCGAATAAAGATGGAGCTTCTTATTTAATTCCTCAAGSTAAAGAGCTCAG 240
 DB 181 GTATCTGATTCGAATAAAGATGGAGCTTCTTATTTAATTCCTCAAGSTAAAGAGCTCAG 240

QY 241 TATAAGGCTTCAACTGATTTAATTCCTTTTACGACAACTACTAATGAGGAGGAACA 300
 DB 241 TATAAGGCTTCAACTGATTTAATTCCTTTTACGACAACTACTAATGAGGAGGAACA 300

QY 301 TATGTAACTAAAAAGATACCTCGCTCAGCAATCAGATTTGGACATGGCTTAATCTATA 360
 DB 301 TATGTAACTAAAAAGATACCTCGCTCAGCAATCAGATTTGGACATGGCTTAATCTATA 360

QY 361 TCAGCTAATACTACACAGTTTCCACTGTTACTGAGTCAAAATATGATGTTGAGGTT 420
 DB 361 TCAGCTAATACTACACAGTTTCCACTGTTACTGAGTCAAAATATGATGTTGAGGTT 420

QY 421 ATTAATGTTTCCCAATATGGATATTTATGTTTCTAGCACTGTTAATATGAGCTGTA 480
 DB 421 ATTAATGTTTCCCAATATGGATATTTATGTTTCTAGCACTGTTAATATGAGCTGTA 480

QY 481 ATTATGTTTACATCTGTAACCTCAAAATGCTACTATTTCATGAAAGAAATACCTGATGCGACA 540
 DB 481 ATTATGTTTACATCTGTAACCTCAAAATGCTACTATTTCATGAAAGAAATACCTGATGCGACA 540

QY 541 TGGGAGATGGTGGTGA AAAAAGTGTAGATCAAAAACGTA CTGGTTGGTGTATACAGTC 600
 DB 541 TGGGAGATGGTGGTGA AAAAAGTGTAGATCAAAAACGTA CTGGTTGGTGTATACAGTC 600

QY 601 AAATATACTATTACTATATAAGATGTCAGTCAATTTATCATGTACAGAAAAGTGTATCAA 660
 DB 601 AAATATACTATTACTATATAAGATGTCAGTCAATTTATCATGTACAGAAAAGTGTATCAA 660

QY 661 TATGTTTAAAGGATACTATGCCATCTGCTTCTGTAGTTGATTTGAACGAAGGCTCTTAT 720
 DB 661 TATGTTTAAAGGATACTATGCCATCTGCTTCTGTAGTTGATTTGAACGAAGGCTCTTAT 720

QY 721 GAAGTAACTATTACTGATGATCAGGGAATATTACAACTTAATCAAGTTCCGAAAAA 780
 DB 721 GAAGTAACTATTACTGATGATCAGGGAATATTACAACTTAATCAAGTTCCGAAAAA 780

QY 781 GCAACTGGGAAGTATTAACCTGTTAGAGGAAAAATAATAATTTACGATTACTATTCCGTGG 840
 DB 781 GCAACTGGGAAGTATTAACCTGTTAGAGGAAAAATAATAATTTACGATTACTATTCCGTGG 840

QY 841 GCAGTACCAATATCTCCAAACCGGAAATACTCAAAATGGAGCTAATGATGACTTTTTTAT 900
 DB 841 GCAGTACCAATATCTCCAAACCGGAAATACTCAAAATGGAGCTAATGATGACTTTTTTAT 900

QY 901 AAGGGATTAATCAATCAACAGTCAGTCACTTATACAGGAGTATTAAGAGTGGAGCTAAACCA 960
 DB 901 AAGGGATTAATCAATCAACAGTCAGTCACTTATACAGGAGTATTAAGAGTGGAGCTAAACCA 960

QY 961 GGTTTCAGCTGATTTACAGGAAAAATACAAAATTCGCGACCACTCAACCCCAATACTAGCAAT 1020
 DB 961 GGTTTCAGCTGATTTACAGGAAAAATACAAAATTCGCGACCACTCAACCCCAATACTAGCAAT 1020

QY 1021 GATGACCCAGCTCAAAAGTAAACAGTGGAGGATGGTCAAAATTTACTATAAAAAAATTTGAT 1080
 DB 1021 GATGACCCAGCTCAAAAGTAAACAGTGGAGGATGGTCAAAATTTACTATAAAAAAATTTGAT 1080

QY 1081 GGTTCACAAAAAGCTTCATTACAAAGGTGCTATATTTGTTTTTAAAGATGCTACGGGTCAA 1140
 DB 1081 GGTTCACAAAAAGCTTCATTACAAAGGTGCTATATTTGTTTTTAAAGATGCTACGGGTCAA 1140

QY 1141 TTTCTAAACTTTAACGATACAAAATAACGTTGAAATGGGGCACAGAGCTAATGCAACAGAA 1200
 DB 1141 TTTCTAAACTTTAACGATACAAAATAACGTTGAAATGGGGCACAGAGCTAATGCAACAGAA 1200

QY 1201 TATACAAAGGAGCAGATGCTATTAATTCATTTACAGGCTTGAAGAAGGTACATCTAT 1260
 DB 1201 TATACAAAGGAGCAGATGCTATTAATTCATTTACAGGCTTGAAGAAGGTACATCTAT 1260

QY 1261 CTAGTTTGAGAAAAAGGCTCCCTTAGGTACAAATTTCTGTAGATAACTCTCAGAAAGTTTAT 1320
 DB 1261 CTAGTTTGAGAAAAAGGCTCCCTTAGGTACAAATTTCTGTAGATAACTCTCAGAAAGTTTAT 1320

QY 1321 TTAGGAGATGGAGCCACTGATACGATAATTCAGATAAACCTTTTAGTTAACCCAACTGTT 1380
 DB 1321 TTAGGAGATGGAGCCACTGATACGATAATTCAGATAAACCTTTTAGTTAACCCAACTGTT 1380

QY 1381 GAAATAACAAGGTACTGATGCTTCCCTTCAACAGGTGGTATTTGGTACAACAATTTTCTAC 1440
 DB 1381 GAAATAACAAGGTACTGATGCTTCCCTTCAACAGGTGGTATTTGGTACAACAATTTTCTAC 1440

QY 1441 ATTATAGGTGCAATTTTAGTAAATAGGACAGGTATCGTCTTGTGCTCGTCTCGTTTA 1500
 DB 1441 ATTATAGGTGCAATTTTAGTAAATAGGACAGGTATCGTCTTGTGCTCGTCTCGTTTA 1500

QY 1501 CGTCTCTAA 1509
 DB 1501 CGTCTCTAA 1509

RESULT 2

ABK11585
 ID ABK11585 standard; DNA; 1509 BP.

XX ABK11585;

XX AC AC

XX 05-JUN-2002 (first entry)

XX S. agalactiae SbpI gene.

DE Extracellular matrix adhesion; Ema; ds; group B streptococcus; GBS;
 KW DNA vaccine; SbpI; Sbp2; Rib; Lmb; C5a-ase; C protein alpha antigen;
 KW neonatal bacterial infection; gene.

XX Streptococcus agalactiae.

XX Key Location/Qualifiers
 FH 1..1509
 FT CDS
 FT /*tag= a
 FT /product= "SbpI"


```

QY 1501 CGTCTTAA 1509
Db 1501 CGTCTTAA 1509

RESULT 3
AAAX20108
ID AAX20108 standard, DNA; 1983 BP.
XX
XX
AC AAX20108;
XX
XX 20-APR-1999 (first entry)
XX
DE Enterococcus faecalis gene EF058.
XX
KW Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
KW detection; attenuation; antigenic; ss.
XX
OS Enterococcus faecalis.
XX
PN WO980554-A2.
XX
PD 12-NOV-1998.
XX
XX 04-MAY-1998; 98WO-US008959.
XX
PR 06-MAY-1997; 97US-0044031P.
PR 16-MAY-1997; 97US-0046655P.
PR 14-NOV-1997; 97US-0066009P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Kunsch CA, Choi GH, Bailey C, Hromockyj A;
XX
XX WPI; 1999-070095/06.
XX P-PSDB; AAY00118.
XX
XX New isolated Enterococcus faecalis polynucleotides - used to develop
XX products for the detection of Enterococcus and for use in vaccines for
XX prevention or attenuation of Enterococcus infection.
XX
XX Claim 1; Page 133; 301pp; English.
XX
XX The present sequence represents a gene isolated from Enterococcus
XX faecalis. The present invention describes genes, proteins and antigenic
XX polypeptides isolated from E. faecalis. The proteins can be used in
XX vaccines for preventing or attenuating an infection caused by a member of
XX the Enterococcus genus in an animal. They can also be used for detecting
XX Enterococcus antibodies in a sample. The nucleotide sequences can be used
XX for detecting Enterococcus nucleic acids. Products from the present
XX invention can also be used for screening compounds to identify agonists
XX and antagonists of E. faecalis protein activity
XX
XX Sequence 1983 BP; 729 A; 330 C; 426 G; 498 T; 0 U; 0 Other;

Query Match 4.3%; Score 64.8; DB 2; Length 1983;
Best Local Similarity 52.1%; Pred. No. 0.00013;
Matches 172; Conservative 0; Mismatches 152; Indels 6; Gaps 1;

QY 1180 ACAGAGCTAATGCAACAGATATACACAGGACGATGGTATATACATTACAGGC 1239
Db 1660 ACAAAAGCTGAAGCAACTACTTTTACACAAACGCGCTGATGGATTAGTTGATATCACAGG 1719
QY 1240 TTGAAGAAGGTACATCTACTATCTAGTTGAGAAAAGGCTCCCTTAGGTTACAATTTGTTA 1299
Db 1720 CTTAAATACGGTACCTATTTTATAGAGAACTGCTAGCTCTCTGATGATGATGCTTTGTTA 1779
QY 1300 GATAACTCTCAGAGGTTATTTAGGAGATGGAGCCACTGATACGACTAATTCAGATAAC 1359
Db 1780 ACAAAATCGGATTGAATTTGGTCAATGAACAAATCATATGCGCAA-----CAGAAAC 1833
QY 1360 CTTTATTAAACCAACTGTTGAAATAACAAAGGTTACTGATGTTGCCCTTCAACAGGTGGT 1419

Db 1834 CTAGTTTCACAGAAAAGTACCAACAAACAAAGGTACCTTACCTTCAACAGGTGGC 1893
QY 1420 ATTGGTACAACAATTTTCTACATTATAGTTCGAATTTTAGTATATAGGACGAGGTATCGTG 1479
Db 1894 AAAGGAATCTACGTTTACTTTAGGAAGTGGCGCAGTCTTGCTACTTATTGCGAGGAGTCTAC 1953
QY 1480 CTTGTTGCTCGTCTGCTGTTTACGTTCTTAA 1509
Db 1954 TTGCTAGACGTAGAAAAGAAATGCTTAA 1983

RESULT 4
ABN98093
ID ABN98093 standard; DNA; 1983 BP.
XX
XX AC ABN98093;
XX
XX 05-AUG-2002 (first entry)
XX
DE E faecalis EF058 gene.
XX
KW Enterococcus; vaccine; gastrointestinal disease; diagnosis; antibiotic;
KW gene; ds.
XX
XX Enterococcus faecalis.
XX
XX US2002045737-A1.
XX
XX 18-APR-2002.
XX
XX 04-MAY-1998; 98US-00071035.
XX
XX 04-MAY-1998; 98US-00071035.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Choi GH, Bailey C, Hromockyj A, Kunsch CA;
XX
XX WPI; 2002-425450/45.
XX P-PSDB; ABP43337.
XX
XX New genes and polypeptides from Enterococcus faecalis, useful as vaccines
XX for preventing, treating or attenuating an infection caused by a member
XX of the Enterococcus genus in an animal, particularly E. faecalis.
XX
XX Claim 1; Page 92; 255pp; English.
XX
XX The present invention provides the protein and coding sequences of a
XX number of polypeptides from Enterococcus faecalis. The proteins can be
XX used as vaccines for preventing or attenuating an infection caused by a
XX member of the Enterococcus genus in an animal, particularly E. faecalis.
XX The polynucleotide is also useful for preventing or treating E. faecalis
XX infection. The present sequence is a coding sequence of the invention
XX
XX Sequence 1983 BP; 729 A; 330 C; 426 G; 498 T; 0 U; 0 Other;

Query Match 4.3%; Score 64.8; DB 6; Length 1983;
Best Local Similarity 52.1%; Pred. No. 0.00013;
Matches 172; Conservative 0; Mismatches 152; Indels 6; Gaps 1;

QY 1180 ACAGAGCTAATGCAACAGATATACACAGGACGATGGTATATACATTACAGGC 1239
Db 1660 ACAAAAGCTGAAGCAACTACTTTTACACAAACGCGCTGATGGATTAGTTGATATCACAGG 1719
QY 1240 TTGAAGAAGGTACATCTACTATCTAGTTGAGAAAAGGCTCCCTTAGGTTACAATTTGTTA 1299
Db 1720 CTTAAATACGGTACCTATTTTATAGAGAACTGCTAGCTCTCTGATGATGATGCTTTGTTA 1779
QY 1300 GATAACTCTCAGAGGTTATTTAGGAGATGGAGCCACTGATACGACTAATTCAGATAAC 1359
Db 1780 ACAAAATCGGATTGAATTTGGTCAATGAACAAATCATATGCGCAA-----CAGAAAC 1833
QY 1360 CTTTATTAAACCAACTGTTGAAATAACAAAGGTTACTGATGTTGCCCTTCAACAGGTGGT 1419

```


Db 13844 ACAATCGGATTGAATTTGTGCTCAATGAACAATCATATGGCACAA-----CAGMAAC 13897
QY 1360 CTTTGTAGTTAACCCAACTGTTGAAATAACAAAGGTACTGAGTTGCCCTTCACAGGTGGT 1419
Db 13898 CTAGTTTACACAGAAAGGTACCAACAAACACAAAGGTACTTACCTTTCACAGGTGGC 13957
QY 1420 ATTGTGACAAACAATTTCTACATTTATAGGTGCAATTTTATAGTAAGGACAGGTATCGTG 1479
Db 13958 AAAGGAATCTAGTTTACTTAGGAAGTGGCGAGTCTTGCTACTATTGTCAGGAGTCTAC 14017
QY 1480 CTTGTTGCTCGTCGTTTACGTTTCTTAA 1509
Db 14018 TTGCTAGACGTAGAAAAGAAAATGCTTAA 14047

RESULT 10

AAA70187
ID AAA70187 standard; DNA; 11922 BP.
XX AC
XX AAA70187;
XX
XX 07-NOV-2000 (first entry)
XX
XX Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:320.
XX
XX Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
XX
XX antimalarial; malaria; protozoacide; infection; insecticide; ds.
XX
XX Plasmodium falciparum.

WO200025728-A2.

11-MAY-2000.

05-NOV-1999; 99WO-US026796.

05-NOV-1998; 98US-0107131P.

(HOFF/) HOFFMAN S.
(CARU/) CARUCCI D.
(GARD/) GARDNER M.
(VENT/) VENTER J C.

Hoffman S, Carucci D, Gardner M, Venter JC;

WPI; 2000-365347/31.

Proteins encoded by chromosome 2 of the human malarial parasite,
Plasmodium falciparum, useful as antimalarial vaccines and in the
diagnosis of P.falciparum infection.

Disclosure; Page 516-519; 577pp; English.

The present invention describes proteins and their fragments (I) encoded
by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
vaccines against P. falciparum infection comprising (I) or (II). (I) and
(II) are useful for the development of vaccines against P. falciparum
infection. (I) and polyclonal antisera or a monoclonal antibody raised to
immunogens comprising the sequences of (I), are useful in the detection
of infection with P. falciparum. Furthermore, (I) (especially when they
are refined or secreted or membrane proteins) can aid the identification
of drugs to treat or prevent P. falciparum infection, or they can be used
to identify drug resistance in P. falciparum. Sequencing of the
Plasmodium chromosome 2 and the subsequent identification of proteins
encoded by it will help to expand our understanding of parasite biology,
a process hampered by the complexity of the parasitic lifecycle, and
provide new targets for vaccine and drug development. Parasite resistance
to drugs and mosquito resistance to insecticides have led to a resurgence
of malaria in many parts of the world, and there is a pressing need for
CC vaccines and new drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352
represent nucleotide and protein sequences given in the present

CC invention, but which are not specifically mentioned within the
CC specification

XX SQ Sequence 11922 BP; 5402 A; 948 C; 1343 G; 4229 T; 0 U; 0 Other;

Query Match 3.9%; Score 59.6; DB 3; Length 11922;
Best Local Similarity 46.3%; Pred. No. 0.0029;
Matches 265; Conservative 0; Mismatches 304; Indels 3; Gaps 2;

QY 111 TCAGATACCTCAAAAGGCGCAACCTATAAAGCATATAAAGTTTTTTCATGATCGAGAAATAGA 170
Db 10038 TCTAGATATTATTAATGACCATACCTTTATGGAGTCTATGAAAGTTGCAACCAATAGGAT 10097
QY 171 TAATGCAATGTATCTGATTGGAATAAAGATGGAGCTTCTTATTATTCCTCAAGGTAA 230
Db 10098 TTATGAAATCCATCTACTATAGGAATTATATAAAGAACCGTACACCTCTTAATGAACAA 10157
QY 231 AGAAGCTGATATAAAGCTTCAACTGATTTTAAATCTCTTTTTTACGACAACTACTATGG 290
Db 10158 AATTTTAAATAGTAAATATGATATTAACAATAGTAATCATTTTTTAAAGAAATAGTATGA 10217
QY 291 AGGAGAACATATGTAACCTAAAAAGATGCTGCTGAGCAATGAGATTCGGACATGGGC 350
Db 10218 AAATGTAACCAACCAACCAATACAAATACAAATACAAATACAAATACAAATAA 10277
QY 351 TAAATCTATATCAGCTAATTAATCAACAGTTTCCACTGTTTACTGAGTCAAAATATGATGG 410
Db 10278 AATTTCTGATACAGATAATCAATATGATGATGATGATGATGATGATGATGATGATGATG 10337
QY 411 TACTGAGTTTAAATGTTTCCCAATATGATGATTAATTAATGTTTCT-AGCACTGTTTAATA 469
Db 10338 TGATGATATAATCATAATGACGATATAATCATAATGATGATGATGATGATGATGATGATG 10397
QY 470 ATGGAGCTGTAATGTTTACATCTGTAA--CTCCAAATGCTACTATTTCATGAGAAAGA 527
Db 10398 ATATAATCATAATGATGATGATTAATCATAATGATGATGATGATGATGATGATGATGAT 10457
QY 528 TACTGATGCGACATGGGAGATGGTGGGAAAACTGTAGATCAAAAAACGCTACTCGGT 587
Db 10458 TCATAATGATGACAGATAATCATAATGATACAGATAATCATAGTATAATTATAATCTCA 10517
QY 588 TGGTATACAGTCAAAATATACTATTACTTATTAAGATGAGTCAATATATCATGATGATGATG 647
Db 10518 TAAATATAAGGAACTTATAAAATTTATCGTATTTCATGATGAGGAAGATATTATTCAAGA 10577
QY 648 AAAAGTGATCAATATGTTATAAAGGATACTA 679
Db 10578 TAATAATTATACAAACGATGATTATGTTAATA 10609

RESULT 11

ACA28118
ID ACA28118 standard; DNA; 1137 BP.

XX AC ACA28118;

XX 19-JUN-2003 (first entry)

XX Prokaryotic essential gene #9775.

DE Antisense; ds; prokaryotic essential gene; cell proliferation;
XX drug design; gene.

XX Clostridium acetobutylicum.

XX WO200277183-A2.

XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
XX (BLIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haseelbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
XX WPI; 2003-0299926/02.
XX P-PSDB; ABU24248.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation or to
PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 14; SEQ ID NO 15988; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 1137 BP; 450 A; 138 C; 217 G; 332 T; 0 U; 0 Other;
Query Match 3.9%; Score 59.2; DB 7; Length 1137;
Best Local Similarity 43.7%; Pred. No. 0.0021;
Matches 262; Conservative 0; Mismatches 338; Indels 0; Gaps 0;
QY 90 TGAGACAGGACAAATACAGTTCAGAGTACTCAAAAGGCGCAACCTATAAGCATATA 149
DB 30 TGGGACAGGCGCAGAGCTATAAGATGGCTCTTTGTTAAAGAACTGGAAGAAATAA 89
QY 150 AGTTTGTGTCAGAAATAGATATGCAATGTATCTGATTCGAATAAGATGAGGCTTC 209
DB 90 AAATTTTCAAGAAAAAGTAGTGTGTACAGCACAGCAGGAGATGCTAGATCAAGTTCT 149
QY 210 TTATTTTAATTCCTCAAGTAAAGAGCTGAGTATAAGCTTCAACTGATTTTAATTCCT 269
DB 150 TAGTTTGTTCATATAGAGCCAGCTTTGATTTGATATATGAAATAAGCAATCTCT 209
QY 270 TTTTACGACAACTACTAATGGAGGGAGAAATATGTAACATAAAAGATACTCGTCAGC 329
DB 210 TACACAGTAACCTCAAGTCTTTTACATGGGTTAAGCGAATATTTCTGCTGAAGACC 269
QY 330 AATGAGATTGCGACATGGGCTGCTAAATCTATATCAGCTAATATACACAGTTTCCACTGT 389

Db 270 AGATATATTTAGTACATGGAGATACAACTACAAAGTTTGCAGCATCACTTGCAGCCTT 329
QY 390 TACTGAGTCAAAATPAATGATGGTACTGAGGTATTAATGTTTCCCAATATGATATTA 449
Db 330 TTATGAAAAGATAGCTATTGGGCTATGAGGCTGGACTTTAGAACTTATGATAAATATT 389
QY 450 TGTTCCTAGCAGCTGTTAATAATGAGAGCTGTAATATGTTTACATCTGTACTCCAATGC 509
Db 390 TCTTTTCTGGAAGATAGAACTTACTGGAGCAATTCGAGATATGCAATTTTTC 449
QY 510 TACTATTCATGAAAAGATATCTATGCGACATGGGAGATGGTGGTGGAAAACTGTAGA 569
Db 450 ACCGACAGTAATATCTTAAATAATATCTTTTAAAGAGAGGTGTTAAAGAAAAATATATT 509
QY 570 TCAAAAAACGTTACTCGGTTGGTGATACAGTCAAAATATCTATTACTTTATAAGATGCAGT 629
Db 510 TATAACTGGAATACTGTGATTGCTATGAAATACACAGTGGATAGTAATATGTTATT 569
QY 630 CAATATCATGGTACAGAAAGAGTGTATCAATATGTTATAAGAGATATGCACTGTC 689
Db 570 TAAAAATGATCAATTAATAATAATTTGACTACAAACATAAAAGATATTATTATGTTAACTGC 629
RESULT 12
ABQ39490/C
ID ABQ39490 standard; DNA; 969 BP.
XX
XX AC ABQ39490;
XX
XX DT 12-JUL-2002 (first entry)
XX
XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 26081.
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX drug; side effect; cancer; central nervous system; cardiovascular;
XX gastrointestinal; respiratory system; single nucleotide polymorphism;
XX SNP; cell differentiation; ds.
XX Homo sapiens.
XX WO200218632-A2.
XX
XX PD 07-MAR-2002.
XX
XX PF 01-SEP-2001; 2001WO-EP010074.
XX
XX PR 01-SEP-2000; 2000DE-01043826.
XX
XX PR 05-SEP-2000; 2000DE-0104543.
XX
XX PA (EPIG-) EPIGENOMICS AG.
XX
XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
XX DR WPI; 2002-371829/40.
XX
XX PT Determining the degree of cytosine methylation in genomic DNA, useful for
XX diagnosis and prognosis, comprises selective hybridization of amplicons
XX from chemically treated DNA.
XX
XX PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
XX CC This invention describes a novel method for determining the degree of
XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX genomic sample of DNA. The sample is treated chemically to convert
XX cytosine (C) but not methylated C, to uracil, then part of the genomic
XX DNA that contains the target C is amplified to form a labeled amplicon.
XX The amplicon is hybridised to two classes, each with at least one member,
XX of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
XX degree of hybridisation to both classes is determined from the label on
XX the amplicon. From the ratio of labels hybridised to the two classes of
XX oligomers, the degree of methylation is calculated. The method is used:
XX (i) for diagnosis and/or prognosis of side effects of therapeutic drugs


```
QY 150 AGTTTGTGATGCGAATATGATGCAATGTATCTGATTCGATTAAGATGAGCTTC 209
Db 181 TACTAATAAAGCAATATAAATAACGAATAAATACTAATAAAGCAATATAAATACGAA 240
QY 210 TTATTTAATCTCCTCAAGGTAAGAGCTGAGTATAAGCTTCAACTGATTTAATTTCTCT 269
Db 241 TAAAATACATAAATAACGAATAAATAACGAATAAATACTAATAAAGCAATATAAATA 300
QY 270 TTTTACGACAACTACTAATGAGGGAGAACATATGTAACTAATAAAGATACCTCGGTACG 329
Db 301 AACGAATAAATAAATACTAATAAAGCAATAAATAAAGCAATAAATACTAATAAAGCA 360
QY 330 AATGAGATTGCGACATGGCTAAATCTATATCAGCTAATACCTACACCACTTTCCACTGT 389
Db 361 TAAAATACGAATAAATAAATACTAATAAAGCAATAAATAAAGCAATAAATACTAATAA 420
QY 390 TACTGAGTCAAAATGATGGTACTGAGGTATTAATGTTTCCCAATATGATGATATTATTA 449
Db 421 AACGAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 480
QY 450 TGTTTCTAGCACTGTTAATAATGAGCTGTAATTAATGGTTTACATCTGTAACTCCAAATGC 509
Db 481 TAAAGCAACGAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 540
QY 510 TACTATTTCATGAAGATACCTGATGCGACATGGGAGATGGTGGTGAATAAATCTGTAGA 569
Db 541 AATACTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 600
QY 570 TCAAAAACCTACTCGGTTGGTGTGATACAGTCAATATATCTATTATTAAGATGCACT 629
Db 601 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 660
QY 630 CAATTAATCATGGTACGAAAGATGTATCAATATGTTTATAAGGATACCTATGCCATCTGC 689
Db 661 GAATACGAATACGAATACAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 720
QY 690 TTCTGTGATTGTTGACGAAGGCTTTATGAGTAACTATTACTATGATGATGATGAGGAA 749
Db 721 TACTAATAAAGCAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 780
QY 750 T---ATTACAACTCTACTCAAGTTTCGGAAGAAAGCAACTGGGAAGTATACCTGTTAGA 806
Db 781 TAAAATACATAAATAAAGCAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 840
QY 807 GGAATAATAATTTACGATTTACTATTCGTTGGGAGCTACCAATATCTCCAAACCGGAAA 866
Db 841 TAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 900
QY 867 TACTCAAAATGGAGCTAATGATGACTTTTATTAAGGGAATAAATAAATAAATAAATAAATA 926
Db 901 AACGAACGAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 960
QY 927 TTATACA 933
Db 961 TAATAAA 967
```

RESULT 14

ABL28164
ID ABL28164 standard; DNA; 2408 BP.

XX ABL28164;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 35965.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical; gene; ds.

XX Drosophila melanogaster.

PN W0200171042-A2.

XX

PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US009231.

XX PR 23-MAR-2000; 2000US-0191637P.

XX PR 11-JUL-2000; 2000US-00614150.

XX PA (PEKE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers BW;

XX WPI; 2001-656860/75.

XX DR New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signaling and cell-cell
XX PT interactions.

XX PS Claim 1; SEQ ID NO 35965; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABU16176-ABU16175), expressed DNA
XX CC sequences (ABU16176-ABU16175) and the encoded proteins (ABU16176-
XX CC ABU16175). The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at fip.wipo.int/pub/published_pct_sequences

XX SQ Sequence 2408 BP; 724 A; 504 C; 437 G; 743 T; 0 U; 0 Other;

Query Match 3.7%; Score 55.4; DB 4; Length 2408;

Best Local Similarity 49.9%; Pred. No. 0.017;

Matches 195; Conservative 0; Mismatches 191; Indels 5; Gaps 2;

QY 20 AATCGCTGTTAGTGGCGAGTTTAGCATTTGGTATGGCTGTATCAC--CAGTTAGCCGAT 77

Db 963 ATTCACTTTTAGCTGTAGCTTTAGCTTATAGCTTTGGCGTCATTAACCAAGATACAGATAC 1022

QY 78 AGCTTTTTCGCGTGAGACAGGGAATTTACAGTTCAAGTACTCAAAAAGCGCAACCTA 137

Db 1023 AGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATAC 1082

QY 138 TAAAGCATATAAAGTTTGTATGATCGAGAAATAGATAATGCAAAATGTATCTGATTGGAATAA 197

Db 1083 AGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATAC 1142

QY 198 AGATGAGCTTCTTATTTAATTTCTCAAGTAAAGAGCTGAGTATAAAGCTTCAACTGA 257

Db 1143 AGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATAC 1199

QY 258 TTTTAAATTTCTTTTACGCAACTACTAATGAGGAGGAGCAATATCTAATCACTAAAAAGA 317

Db 1200 TACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATAC 1259

QY 318 TACTGCGTTCAGCAAAATGAGATTCGCGACATGGGCTAAATCTATATCAGCTAATACACCC 377

Db 1260 TACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATAC 1319

QY 378 AGTTTCCACTGTTTACTGAGTCAAAATATGAT 408

Db 1320 TACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATAC 1350

RESULT 15

ABL56203/c

ID ABL56203 standard; DNA; 32392 BP.

XX ABL56203;

XX 01-JUL-2002 (first entry)

XX AmEPV genome fragment#5.
 XX AmEPV; gene therapy; viral vector; chromosome mapping; gene mapping;
 KW genetic deficiency disorder; ds.
 XX Amsacta moorei entomopoxvirus.
 XX WO200212526-A2.
 XX 14-FEB-2002.
 XX 10-AUG-2001; 2001WO-US025287.
 XX 10-AUG-2000; 2000US-0224479P.
 XX 14-SEP-2000; 2000US-00662254.
 XX (UYFL) UNIV FLORIDA.
 XX Moyer RW, Li Y, Bawden AL;
 XX WPI; 2002-227161/28.
 XX Novel recombinant entomopox virus vector useful for delivering
 PT polynucleotide encoding protein to vertebrate cell, comprises
 PT polynucleotide encoding protein operably linked with heterologous
 PT promoter sequence.
 XX Disclosure; Page 226-242; 326pp; English.
 XX The invention relates to a recombinant entomopox virus (EPV) vector,
 CC comprising a polynucleotide encoding a protein operably linked with a
 CC heterologous promoter sequence. The invention also concerns methods for
 CC providing gene therapy for genetic deficiency disorders. Vectors of the
 CC invention are useful for delivering a polynucleotide encoding a protein
 CC to a vertebrate cell preferably a mammalian cell, such as a human cell.
 CC The vector is introduced into the vertebrate cell by infection in a viral
 CC particle, or by transfection, transduction, or injection either in vitro
 CC or in vivo. The vector is useful for the delivery and expression of
 CC biologically useful proteins in gene therapy protocols, and for
 CC delivering large DNA segments for engineering of vertebrate cells.
 CC Polynucleotides of the invention have applications in techniques such as
 CC their use as insertion sites for foreign genes of interest, hybridisation
 CC probes, for chromosome and gene mapping, in PCR technologies, and in the
 CC production of sense or antisense nucleic acids. Vectors of the invention
 CC provide for stable integration and expression of heterologous DNA in host
 CC cells, and are adapted for accepting large heterologous polynucleotide
 CC inserts which can be delivered in an infected or transformed cell and
 CC expressed in a stable fraction. The current sequence represents a
 CC fragment of the genome of the genus B entomopoxvirus from amsacta moorei
 CC (AmEPV)
 XX
 SQ Sequence 32392 BP; 13748 A; 2577 C; 2550 G; 13517 T; 0 U; 0 Other;
 Query Match 3.7%; Score 55.2; DB 6; Length 32392;
 Best Local Similarity 44.0%; Pred. No. 0.034;
 Matches 329; Conservative 0; Mismatches 413; Indels 6; Gaps 2;
 QY 479 TAATTATGGTTACATCTGTAACTCCAAATCGCTCTATTTCATGAAAGAAATACGTGCGGA 538
 DB 1602 TAATAATATAAATGGATATATTACTAATTCGCTCTAACTATCATATTTTCAGATCATA 1543
 QY 539 CATGGGGAGATGGTGGGAAAGAACTGTAGATCAAAAAGCTACTCGGTGGTGATACAG 598
 DB 1542 ACATATATGTTGATGATGAGAAAATAGAGATTTTAATAAATAATATCTCGGTGTTAA 1483
 QY 599 TCAATAATATCTATTATTAAGAAATGCAATGCTCAATATCATGTCAGAGAAAAGTGTATC 658
 DB 1482 TTAATAATTAATAATA---ACAATAATACTAAAAAAATATACTACTCAATAATAGATAATT 1426
 QY 659 AATATGTTATAAGGATACATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 718
 DB 1425 ATAATGAATAGAAAAAATAAATTAATGAATAAGTGTGATGATGATGATGATGATGATGAT 1369

QY 719 ATGAAGTAACCTATTACTGATGGATCAGGAATATTCAACTCTAACTCAAGGTTCCGAAA 778
 DB 1368 ATTGCACAGATCTAACTAATAAATAAGCAATATTATTCAAAATCAGTATATAGTAAAA 1309
 QY 779 AAGCACTGGGAAGTAGTAACCTGTTAGAGGAAAATAATAATTTTCAAGTACTATTCCGT 838
 DB 1308 CAGAAGAAGGGAGTAAATCTGTTATTAACTGATATCTGATAATAATTAATAAATAGTT 1249
 QY 839 GGGAGCTACCAATACCTCCAAACCGGAAATACTCAAAATGGAGCTAATGATGACTTTTTT 898
 DB 1248 TAGCAGAAGATATATCTTTTAAAAAGAGGATGATAAAGAAAGATTATAATATAGATTAT 1189
 QY 899 ATAAGGGAATAAATACAAATCACAGTCACTTATACAGAGCTATTAAAGAGTGGAGCTAAAC 958
 DB 1188 ATGAGAAAATAAATAATGTAATAATACAAATAGAAATAAATAATATATATCAAAAAATTAT 1129
 QY 959 CAGGTTGAGCTGATTTTACCAGAAAATACAAACATTCGACCATCAACCCCAATPACTAGCA 1018
 DB 1128 TATCTTTTGATTATAATAATAAAAAATAAATAATAATAGTGTGCTGTAATAATTTGATGGA 1069
 QY 1019 ATGATGACCCAGGTCAAAAAGTAAACAGTGGAGGATGCTCAATTTACTATAAAAAATTG 1078
 DB 1068 TTATTAAACGATGATTAAATTTGAAAGCAAAATTCGAAAGAAAAGACTATTTCAAATACATTC 1009
 QY 1079 ATGGTTCCCAAAAGCTTCATTACAAAGTGTCTATATTGTTTTTAAAGAATGCTACGGGTC 1138
 DB 1008 CTATATATGAAAAAGTTCAATTAGAATAATATTATTTTATTGATTTATAAAGAGTA 949
 QY 1139 AATTTCTAACTTTAAAGTACAAATAACGTTGAATGGGGCACAGAGCTAATGCAACAG 1198
 DB 948 AATTAATTCAACATTTATGATAACGAAATAGATGTAATAGATTATATATACCGACGAACAAT 889
 QY 1199 AATATACAACAGGACGAGATGGTATAAT 1226
 DB 888 TATTTCTGAATAATGTAATAATAATAAT 861

Search completed: September 13, 2004, 02:13:17
 Job time : 608.354 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 2004, 01:03:46 ; Search time 5983.67 Seconds
(without alignments)
10930.517 Million cell updates/sec

Title: US-10-009-254-1
Perfect score: 1509
Sequence: 1 atgaaaagaagaatgattca.....gtcgctgttaacgtttcttaa 1509

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_man.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1509	100.0	1509	1	AF485279	AP485279 Streptococcus
2	1509	100.0	1509	6	AX476887	AX476887 Sequence
3	92.8	6.1	216959	10	AC116998	AC116998 Mus muscu
4	89.6	5.9	171104	2	BX649545	BX649545 Danio rer
5	87.2	5.8	170221	10	AC127421	AC127421 Mus muscu
6	86.4	5.7	213544	2	BX510939	BX510939 Danio rer
c 7	86	5.7	116807	2	BX890614	BX890614 Danio rer
c 8	85.2	5.6	146275	2	AL935272	AL935272 Danio rer
c 9	84	5.6	198743	2	BX530070	BX530070 Danio rer
c 10	83.4	5.5	155204	2	AC007926	AC007926 Trypanoso
c 11	82.4	5.5	94534	5	AL929250	AL929250 Zebrafish
c 12	80.6	5.3	110000	2	PFMAL13.07	Continuation (8 of
c 13	80.4	5.3	225581	2	BX537105	BX537105 Danio rer
c 14	80	5.3	10115	5	AF397467	AF397467 Ictalurus
c 15	79.8	5.3	308092	3	AY333070	AY333070 Drosophil
c 16	78.6	5.2	224391	2	AC137124	AC137124 Mus muscu
c 17	78	5.2	164399	3	PFMAL3P6	298551 Plasmodium
c 18	77.8	5.2	154160	2	BX571709	BX571709 Danio rer
c 19	77.8	5.2	177595	2	AL953867	AL953867 Danio rer
c 20	77.4	5.1	200634	5	BX004990	BX004990 Zebrafish
c 21	77.2	5.1	176010	5	AL935062	AL935062 Zebrafish
c 22	75.6	5.0	151341	5	AL929536	AL929536 Zebrafish
c 23	75.4	5.0	22398	5	FRU271723	AJ271723 Fugu rubr
c 24	75.2	5.0	245232	2	BX088589	BX088589 Danio rer
c 25	75	5.0	183357	2	BX569779	BX569779 Danio rer
c 26	75	5.0	183357	2	BX569779	BX569779 Danio rer
c 27	74	4.9	199698	2	BX004991	BX004991 Danio rer
c 28	73.8	4.9	250029	3	AB014820	AB014820 Plasmodiu
c 29	73	4.8	666	8	AF413050S2	AF413051 Zea mays
c 30	73	4.8	199347	2	BX321884	BX321884 Danio rer
c 31	72.8	4.8	146907	2	AC141031	AC141031 Rattus no
c 32	72.6	4.8	173134	5	AL935282	AL935282 Zebrafish
c 33	72.4	4.8	157033	2	BX323881	BX323881 Danio rer
c 34	72.4	4.8	179553	2	AC024253	AC024253 Homo sapi
c 35	72.2	4.8	153751	3	AC116551	AC116551 Dictyoste
c 36	72.2	4.8	180668	2	AC020857	AC020857 Mus muscu
c 37	70.6	4.7	146415	5	AL929469	AL929469 Zebrafish
c 38	70.6	4.7	212722	2	AC098909	AC098909 Rattus no
c 39	70.2	4.7	2167	3	PRSTARPA	Z30339 P. reichenow
c 40	69.6	4.6	202182	2	AC120639	AC120639 Rattus no
c 41	69.6	4.6	274060	2	AC095184	AC095184 Rattus no
c 42	69.4	4.6	203271	2	BX322566	BX322566 Danio rer
c 43	69.2	4.6	11534	3	PFAARPIPR	Y08926 P. faiciparu
c 44	69	4.6	178664	2	BX511115	BX511115 Danio rer
c 45	69	4.6	180023	2	BX897685	BX897685 Danio rer

ALIGNMENTS

RESULT 1
AP485279 1509 bp DNA linear BCT 25-NOV-2003
LOCUS Streptococcus agalactiae Spbl (spbl) gene, complete cds.
DEFINITION Streptococcus agalactiae
ACCESSION AP485279
VERSION AP485279.1 GI:28628942
KEYWORDS Streptococcus agalactiae
SOURCE Streptococcus agalactiae
ORGANISM Streptococcus agalactiae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
REFERENCE 1 (bases 1 to 1509)
AUTHORS Adderson, E.E., Takahashi, S., Wang, Y., Armstrong, J., Miller, D.V. and
Bohnsack, J.F.
TITLE Subtractive Hybridization Identifies a Novel Predicted Protein

Mediating Epithelial Cell Invasion by Virulent Serotype III Group B
Streptococcus agalactiae
 Infect. Immun. 71 (12), 6857-6863 (2003)

JOURNAL
 PUBMED
 14638773

REFERENCE
 2 (bases 1 to 1509)

AUTHORS
 Adderson, E. and Bohnsack, J. F.

TITLE
 Direct Submission
 Submitted (19-FEB-2002) Infectious Diseases, St. Jude Children's
 Research Hospital, 332 N. Lauderdale St., Memphis, TN 38105, USA

JOURNAL

FEATURES

source

1..1509
 /organism="Streptococcus agalactiae"
 /mol_type="genomic DNA"

/strain="874391"

/serotype="RDP type III-3"

/db_xref="taxon:1311"

1..1509

/gene="spbl"

1..1509

/gene="spbl"

/note="adhesin/invasin; surface protein of group B"

streptococcus 1"

/codon_start=1

/transl_table=11

/product="Spbl"

/protein_id="AA049409.1"

/db_xref="GI:28628943"

/translation="MKKMIQSLVSLAFGMVSPVTPPIAFRAETGTITVDQPKGA
 TYKAVKVFDAISIDNVDNKGASVLIPOGKEAEYKASTDQNSLFTTYINGGRIV
 TKKDTASNEIATWAKISANTPVSIVIESNNDGTEVINVSQDFTTYINGGRIV
 IYVSVTPNATHEKNIDATWGGGGTVDQKTSVGDVVKYITTYKNAVNYHGTKEV
 YQYVKOTMPSASVDLNEGSEYVITDGSNITLLTQSEKATGKNLLENNFTI
 TTPWAATNPTGNTQNGANDDFKIGINTITVTVGLKSGAKPGSLPENTNIAT
 NPNTSNDPQKQVTRDQOIITKIDGSTRASIOGAIFVLKNAFGQFLNFDNTNVEM
 GTEANATEYTTGAGLITITGLKEGTVLVEKKAPLGVNLLDSQKVLGDGADFTN
 SDNLLNPTVENVKGTLPSTGGIGTIFPIIIGALLVIGAGVLVARRLRIS"

ORIGIN

Query Match 100.0%; Score 1509; DB 1; Length 1509;

Best Local Similarity 100.0%; Pred. No. 3e-238;

Matches 1509; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGAAAAGAAATGATTCAATCGCTGTAGTGGCGAGTTTAGCATTTGGTATGGCTGTA 60

Db 1 ATGAAAAGAAATGATTCAATCGCTGTAGTGGCGAGTTTAGCATTTGGTATGGCTGTA 60

Qy 61 TCACAGATTACCGGATAGCTTTTCCGCTGAGACAGGCAATTTACAGTTCAAGTACT 120

Db 61 TCACAGATTACCGGATAGCTTTTCCGCTGAGACAGGCAATTTACAGTTCAAGTACT 120

Qy 121 CAAAAGCGCAACTATAAGCATAAAGTTTTTGTATGATGAGAAATAGATAATGCAAAAT 180

Db 121 CAAAAGCGCAACTATAAGCATAAAGTTTTTGTATGATGAGAAATAGATAATGCAAAAT 180

Qy 181 GTATCTGATTCGAATAAGATGGAGCTCTTATTATTCTTCTCAAGTAAAGAGCTGAG 240

Db 181 GTATCTGATTCGAATAAGATGGAGCTCTTATTATTCTTCTCAAGTAAAGAGCTGAG 240

Qy 241 TATAAAGCTTCAACTGATTTTAAATTCCTTTTACGACCACTACTAATGAGGAGGAACA 300

Db 241 TATAAAGCTTCAACTGATTTTAAATTCCTTTTACGACCACTACTAATGAGGAGGAACA 300

Qy 301 TATGTAATAAAGATACGTGCTCAGCAAAATAGATGCGACATGGGCTAAATCTATA 360

Db 301 TATGTAATAAAGATACGTGCTCAGCAAAATAGATGCGACATGGGCTAAATCTATA 360

Qy 361 TCAGCTAATCTACACAGTTTCCAGTGTACTGAGTCAATATGATGCTAGGTT 420

Db 361 TCAGCTAATCTACACAGTTTCCAGTGTACTGAGTCAATATGATGCTAGGTT 420

Qy 421 ATTAATGTTTCCCAATATGATATATTATGTTTCTAGCACTGTTAATAATGGAGCTGTA 480

Db 421 ATTAATGTTTCCCAATATGATATATTATGTTTCTAGCACTGTTAATAATGGAGCTGTA 480

Qy 481 ATTATGTTACATCTGTAACTCCAAATGCTACTATTTCATGAAGAATACACTGATCGACA 540

Db 481 ATTATGTTACATCTGTAACTCCAAATGCTACTATTTCATGAAGAATACACTGATCGACA 540

Qy 541 TGGGAGATGGTGGGAAAACCTGTAGATCAAAAAACGTACTCGGTTGGTGATACAGTC 600

Db 541 TGGGAGATGGTGGGAAAACCTGTAGATCAAAAAACGTACTCGGTTGGTGATACAGTC 600

Qy 601 AAATATATCTATTACTTATAGAATGCACTCAATTAATCATGGTACAGAAAAAGTGATCAA 660

Db 601 AAATATATCTATTACTTATAGAATGCACTCAATTAATCATGGTACAGAAAAAGTGATCAA 660

Qy 661 TATGTTATAAAGGATACCTATGCCATCTGCTTCTGTAGTTCATTTGACGAGGCTCTAT 720

Db 661 TATGTTATAAAGGATACCTATGCCATCTGCTTCTGTAGTTCATTTGACGAGGCTCTAT 720

Qy 721 GAAGTAACTATTACTGTAGGATCAGGGAATPATTCAACTCTAACTCAAGGTTCCGAAAAA 780

Db 721 GAAGTAACTATTACTGTAGGATCAGGGAATPATTCAACTCTAACTCAAGGTTCCGAAAAA 780

Qy 781 GCAACTGGGAAGTATACCTGTTAGAGGAAAAATAAATTTTCAGATTACTATTCCTGG 840

Db 781 GCAACTGGGAAGTATACCTGTTAGAGGAAAAATAAATTTTCAGATTACTATTCCTGG 840

Qy 841 GCAGCTACCAATACTCCCAACCGGAAATACTCAAAATGGAGCTAAATGATGACTTTTTTAT 900

Db 841 GCAGCTACCAATACTCCCAACCGGAAATACTCAAAATGGAGCTAAATGATGACTTTTTTAT 900

Qy 901 AAGGGAATAAATAACAATCAAGTCACTTATACAGGAGTATTAAGAGTGGAGCTAAACCA 960

Db 901 AAGGGAATAAATAACAATCAAGTCACTTATACAGGAGTATTAAGAGTGGAGCTAAACCA 960

Qy 961 GGTTCAGCTGATTTCACAGAAATACAAACATTCGACCATCAACCCCAATACAGCAAT 1020

Db 961 GGTTCAGCTGATTTCACAGAAATACAAACATTCGACCATCAACCCCAATACAGCAAT 1020

Qy 1021 GATGACCCAGGTCAAAAAGTAAACAGTGAGGAGTGGTCAAAATTAATAAAAAAATTTGAT 1080

Db 1021 GATGACCCAGGTCAAAAAGTAAACAGTGAGGAGTGGTCAAAATTAATAAAAAAATTTGAT 1080

Qy 1081 GGTTCACAAAAGCTTCATTACAGGTGCTATATTTGTTTAAAGAAATGCTACGGGTCAA 1140

Db 1081 GGTTCACAAAAGCTTCATTACAGGTGCTATATTTGTTTAAAGAAATGCTACGGGTCAA 1140

Qy 1141 TTTCTAAAACCTTTAACCATACAAATACGTTGAATGGGCGACAGAACTAATGCAACAGAA 1200

Db 1141 TTTCTAAAACCTTTAACCATACAAATACGTTGAATGGGCGACAGAACTAATGCAACAGAA 1200

Qy 1201 TATACACAGGAGCAGATGGTATTAATTAACCATTAACAGGCTTGAAAAGAGGTACATACTAT 1260

Db 1201 TATACACAGGAGCAGATGGTATTAATTAACCATTAACAGGCTTGAAAAGAGGTACATACTAT 1260

Qy 1261 CTAGTTTGAGAAAAGGCTCCCTTAGGTTACAAATTTGTTAGATAAATCTCAGAAAGGTTAT 1320

Db 1261 CTAGTTTGAGAAAAGGCTCCCTTAGGTTACAAATTTGTTAGATAAATCTCAGAAAGGTTAT 1320

Qy 1321 TTAGGAGATGGAGCCACTGATACGCTAATTCAGATAACCTTTTAGTTAACCCCACTGTT 1380

Db 1321 TTAGGAGATGGAGCCACTGATACGCTAATTCAGATAACCTTTTAGTTAACCCCACTGTT 1380

Qy 1381 GAAAAATAACAAAGGTACTGAGTTGCCCTTCAACAGGTTGTTGGTACAAACAATTTTCTAC 1440

Db 1381 GAAAAATAACAAAGGTACTGAGTTGCCCTTCAACAGGTTGTTGGTACAAACAATTTTCTAC 1440

Qy 1441 ATTATAGTGCATTTTAGTAATAGAGCAGGTATCGTCTGTTGCTCGTCTGCTGTTA 1500

Db 1441 ATTATAGTGCATTTTAGTAATAGAGCAGGTATCGTCTGTTGCTCGTCTGCTGTTA 1500

Qy 1501 CGTTCCTTAA 1509

Db 1501 CGTTCCTTAA 1509

RESULT 2
AX476887
LOCUS AX476887 1509 bp DNA linear PAT 12-AUG-2002
DEFINITION Sequence 15 from Patent WO0212294.
ACCESSION AX476887
VERSION AX476887.1 GI:22216151
KEYWORDS
SOURCE Streptococcus agalactiae
ORGANISM Streptococcus agalactiae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1
AUTHORS Adderson, E. and Bohnsack, J.
TITLE Group b streptococcus polypeptides nucleic acids and therapeutic
compositions and vaccines thereof
JOURNAL Patent: WO 0212294-A 15 14-FEB-2002;
ST. JUDE CHILDREN'S RESEARCH HOSPITAL (US) ; University of Utah
Research Foundation (US)
FEATURES
source Location/Qualifiers
1..1509
/organism="Streptococcus agalactiae"
/mol_type="unassigned DNA"
/db_xref="taxon:1311"
1..1509
/notes="unassigned protein product"
/codon_start=1
/transl_table=11
/protein_id="CAD43673.1"
/db_xref="GI:22216152"
/db_xref="RENTREMBL:CAD43673"
/translation="MKKMIQSLVSLAFGMVSPVPIAFNAETGTTIVQDTQGA
TKKDYVFAEDIDNAVSDNSKDGASYLIPQGEABYKASTDFNSLFTTTNGGRYV
TKKDTASABEIAFWAKSISANTTPSVTTESSNDGTEVINQYGYYSSTVNGAV
IMVTSVPNATHEKNTDAWGGGKTDVQKTSVGTVKYITVKNNAVYHGTETK
YQVVIDTPWASVDLNEGSYEVTITDGSNITLTQSEKATGVNLENNFTI
TIPWANTPTQNGANDDFYKGINITITTYTGVLSGAKPGSADLPENTNIATI
NPTSDNDPQKVTYDQGITIKKIDGTSKASLQGAIFVLKNATGGQFLNFDNNVEM
GTENATEYTGADGIIITGLKEGTYLVKPKAPIGLYNLLDNSQKVLGDGADFTN
SDLLNVPTVNNKGTSLPSTGGIGITTFIYIIGAILVIGAGIVLVARRRLRS"
ORIGIN
Query Match 100.0%; Score 1509; DB 6; Length 1509;
Best Local Similarity 100.0%; Pred. No. 3e-239;
Matches 1509; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGAAAAGAAAATGATTCAATCGCTGTAGTGGCGAGTTTAGCATTTGGTATGGCTGTA 60
DB 1 ATGAAAAGAAAATGATTCAATCGCTGTAGTGGCGAGTTTAGCATTTGGTATGGCTGTA 60
QY 61 TCACCAAGTTACGCCGATAGCTTTTGGCGCTGAGACAGGGA CAATTACAGTTCAAGATACT 120
DB 61 TCACCAAGTTACGCCGATAGCTTTTGGCGCTGAGACAGGGA CAATTACAGTTCAAGATACT 120
QY 121 CAAAAGGCGCAACCTATAAGCATATAAAGTTTTCATCGAGAAATAGATAATGCAAT 180
DB 121 CAAAAGGCGCAACCTATAAGCATATAAAGTTTTCATCGAGAAATAGATAATGCAAT 180
QY 181 GTATCTGATTTCGAATAAAGATGAGCTTCTATTATTAATTCCTCAAGTAAAGAGCTGAG 240
DB 181 GTATCTGATTTCGAATAAAGATGAGCTTCTATTATTAATTCCTCAAGTAAAGAGCTGAG 240
QY 241 TATAAAGCTTCAACTGATTTTAAATTCCTTTTTCGACAACTACTAATGAGGAGGAACA 300
DB 241 TATAAAGCTTCAACTGATTTTAAATTCCTTTTTCGACAACTACTAATGAGGAGGAACA 300
QY 301 TATGTAACCTAAAAGATACCTGGTTCAGCAAAATGAGATTTCGACATGGGCTTAATCTATA 360
DB 301 TATGTAACCTAAAAGATACCTGGTTCAGCAAAATGAGATTTCGACATGGGCTTAATCTATA 360
QY 361 TCAGCTAATACTACACCAAGTTTCCACTGTTACTGAGTCAAAATGATGTTACTGAGGTT 420
DB 361 TCAGCTAATACTACACCAAGTTTCCACTGTTACTGAGTCAAAATGATGTTACTGAGGTT 420

QY 421 ATTAATGTTTCCCAATATGGATATATTATGTTTCTAGCACTGTTAATATGAGCTGTA 480
DB 421 ATTAATGTTTCCCAATATGGATATATTATGTTTCTAGCACTGTTAATATGAGCTGTA 480
QY 481 ATTATGTTTACATCTCTGTAATCTCAAAATGCTACTATTTCATGAAAAGAAATCTGATGCGACA 540
DB 481 ATTATGTTTACATCTCTGTAATCTCAAAATGCTACTATTTCATGAAAAGAAATCTGATGCGACA 540
QY 541 TGGGGAGATGGTGGTGGAAAAAAGCTAGATCAAAAAAAGCTAGCTCGGTGGTGTATACAGTC 600
DB 541 TGGGGAGATGGTGGTGGAAAAAAGCTAGATCAAAAAAAGCTAGCTCGGTGGTGTATACAGTC 600
QY 601 AATATATCTATTACTTTATAGAAATGAGTCAATTTATCATGGTACAGAAAAGTGTATCAA 660
DB 601 AATATATCTATTACTTTATAGAAATGAGTCAATTTATCATGGTACAGAAAAGTGTATCAA 660
QY 661 TATGTTTATAAAGGATACTATGCTCTCTGTAGTGTGATTTGAACCAAGAGGCTCTTAT 720
DB 661 TATGTTTATAAAGGATACTATGCTCTCTGTAGTGTGATTTGAACCAAGAGGCTCTTAT 720
QY 721 GAAGTAACTATTACTGATGATCAGGGAATATTAACAATCTAACTCAAGTTCGGA AAAA 780
DB 721 GAAGTAACTATTACTGATGATCAGGGAATATTAACAATCTAACTCAAGTTCGGA AAAA 780
QY 781 GCAACTGGGAGATTAACCTGTTAGAGGAAAATAAATAATTTACAGTACTATTTCGCTGG 840
DB 781 GCAACTGGGAGATTAACCTGTTAGAGGAAAATAAATAATTTACAGTACTATTTCGCTGG 840
QY 841 GCAGCTTACCAATACTCTCAACCGGAAATATCTCAAAATGGAGCTAATGATGACTTTTTTAT 900
DB 841 GCAGCTTACCAATACTCTCAACCGGAAATATCTCAAAATGGAGCTAATGATGACTTTTTTAT 900
QY 901 AAGGGAATAAATAACAATCAAGTCACTTATACAGGAGTATTAAGAGTGGAGCTAAACCA 960
DB 901 AAGGGAATAAATAACAATCAAGTCACTTATACAGGAGTATTAAGAGTGGAGCTAAACCA 960
QY 961 GGTTCAGCTGATTACAGAAAATAACAATTCGCAACCATCAACCCCAATACATAGCAAT 1020
DB 961 GGTTCAGCTGATTACAGAAAATAACAATTCGCAACCATCAACCCCAATACATAGCAAT 1020
QY 1021 GATGACCCAGGTCAAAAAGTAAACAGTGGAGGATGGTCAAAATTAATAAAAAAATTTGAT 1080
DB 1021 GATGACCCAGGTCAAAAAGTAAACAGTGGAGGATGGTCAAAATTAATAAAAAAATTTGAT 1080
QY 1081 GGTTCACAAAAAGCTTCATTATCAAGGTGCTATATTTGTTTTTAAAGAAATGCTACGGGTCAA 1140
DB 1081 GGTTCACAAAAAGCTTCATTATCAAGGTGCTATATTTGTTTTTAAAGAAATGCTACGGGTCAA 1140
QY 1141 TTTCTAAACCTTTAAACGATACAAATAACGTTGAATGGGGCACAGAGCTAATGCAACAGAA 1200
DB 1141 TTTCTAAACCTTTAAACGATACAAATAACGTTGAATGGGGCACAGAGCTAATGCAACAGAA 1200
QY 1201 TATACAAACAGGACAGATGCTATAATTCATTCAGGCTTGAAAGAGGTACATACAT 1260
DB 1201 TATACAAACAGGACAGATGCTATAATTCATTCAGGCTTGAAAGAGGTACATACAT 1260
QY 1261 CTAGTTGAGAAAAGGCTCCCTTTAGGTTACAAATTTGTTAGATACTCTCAGAGAGTTATT 1320
DB 1261 CTAGTTGAGAAAAGGCTCCCTTTAGGTTACAAATTTGTTAGATACTCTCAGAGAGTTATT 1320
QY 1321 TTAGGAGATGGAGCCACTGATACGACTAATTCAGATAACCTTTTAGTTAACCCCACTGTT 1380
DB 1321 TTAGGAGATGGAGCCACTGATACGACTAATTCAGATAACCTTTTAGTTAACCCCACTGTT 1380
QY 1381 GAAAAATAACAAAGGTACTGAGTTGCCCTTCAACAGGTGGTATTCGTACAAACAAATTTCTAC 1440
DB 1381 GAAAAATAACAAAGGTACTGAGTTGCCCTTCAACAGGTGGTATTCGTACAAACAAATTTCTAC 1440
QY 1441 ATTATAGGTGCAATTTTATGTAATAGGAGCAGGATCGTGTCTGTTGCTCGTCGCTGTTA 1500
DB 1441 ATTATAGGTGCAATTTTATGTAATAGGAGCAGGATCGTGTCTGTTGCTCGTCGCTGTTA 1500

Qy	1501 CGTTCCTAA 1509
Db	1501 CGTTCCTAA 1509
RESULT 3	
AC116998	
LOCUS	AC116998 216959 bp DNA linear ROD 08-NOV-2003
DEFINITION	Mus musculus BAC clone RP23-16F6 from 18, complete sequence.
ACCESSION	AC116998
VERSION	AC116998.3 GI:24080765
KEYWORDS	HTG.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 216959) Iwakura, A., Kozlowicz, A., Spalding, L. and Mangiapanello, L. The sequence of Mus musculus BAC clone RP23-16F6 Unpublished (2001) 2 (bases 1 to 216959) Wilson, R. Sequencing of Mus musculus Unpublished (2001) 3 (bases 1 to 216959) McPherson, J.D. and Waterston, R.H. Direct Submission Submitted (05-APR-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA 4 (bases 1 to 216959) McPherson, J.D. and Waterston, R.H. Direct Submission Submitted (28-MAY-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA 5 (bases 1 to 216959) McPherson, J.D. and Waterston, R.H. Direct Submission Submitted (17-OCT-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA 6 (bases 1 to 216959) Wilson, R. Direct Submission Submitted (08-NOV-2003) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA On Oct 17, 2002 this sequence version replaced gi:21218578. ----- Genome Center Center: Washington University Genome Sequencing Center Center code: WUGSC Web site: http://genome.wustl.edu Contact: submissions@wustl.wustl.edu ----- Summary Statistics ----- Center project name: M_BA0016F06
COMMENT	NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions. This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest. MAPPING INFORMATION: Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu

TITLE
JOURNAL
Submitted (15-JUL-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 170221)
McPherson, J.D. and Waterston, R.H.
Direct Submission
TITLE
JOURNAL
Submitted (18-JAN-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
5 (bases 1 to 170221)
McPherson, J.D. and Waterston, R.H.
Direct Submission
TITLE
JOURNAL
Submitted (04-FEB-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
6 (bases 1 to 170221)
Wilson, R.
Direct Submission
TITLE
JOURNAL
Submitted (13-NOV-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Feb 4, 2003 this sequence version replaced gi:27802050.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu>
Contact: submissions@watson.wustl.edu
----- Summary Statistics

Center project name: M_BB0314F01

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
<http://genome.wustl.edu>

SOURCE INFORMATION:
The RPCI-24 BAC Library has been constructed by Pieter de Jong and
coworkers (<http://www.chori.org>) from male C57BL/6J mouse spleen
and/or brain genomic DNA. The clone and detailed information can be
obtained from Pieter de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone. This clone is
overlapped by AC116998.

FEATURES
source
1. 170221
Location/Qualifiers
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="18"
/map="18"
/clone="RP24-314F1"
/clone_lib="RPCI-24"
728. 831
/rpt_family="L1"
2517. 2657
/rpt_family="Alu"
9517. 9652
/rpt_family="ERVK"
9669. 9781
/rpt_family="B4"

repeat_region
9857. 10052
/rpt_family="RMER1A"
11379. 11456
/rpt_family="ERV1"
11502. 11577
/rpt_family="ERV1"
11589. 11670
/rpt_family="ERV1"
11696. 11747
/rpt_family="ERV1"
14141. 14466
/rpt_family="MaLR"
15137. 15194
/rpt_family="MaLR"
20668. 20925
/rpt_family="B4"
21376. 21458
/rpt_family="ERV1"
22395. 22448
/rpt_family="Alu"
23930. 24043
/rpt_family="MIR"
24183. 24263
/rpt_family="Alu"
25105. 25261
/rpt_family="Alu"
25263. 26151
/rpt_family="ERVK"
26170. 26253
/rpt_family="Alu"
26483. 26644
/rpt_family="MIR"
27338. 27756
/rpt_family="MaLR"
28247. 28690
/rpt_family="MaLR"
29190. 29265
/rpt_family="ID"
30206. 30535
/rpt_family="ERVK"
30685. 31038
/rpt_family="MaLR"
31561. 31704
/rpt_family="Alu"
32067. 32214
/rpt_family="Alu"
32231. 32489
/rpt_family="L1"
32504. 32715
/rpt_family="B2"
33729. 34156
/rpt_family="L1"
34679. 34935
/rpt_family="B4"
35007. 35111
/rpt_family="L1"
36037. 36186
/rpt_family="L1"
36329. 36437
/rpt_family="Alu"
36828. 36900
/rpt_family="Alu"
37058. 37192
/rpt_family="Alu"
37599. 37812
/rpt_family="B2"
37961. 38151
/rpt_family="B2"
38477. 38557
/rpt_family="ERV1"
39224. 39320
/rpt_family="B2"
39260. 39576

* 73740 96363: contig of 22624 bp in length
 * 96364 96463: gap of 100 bp
 * 96464 99575: contig of 3112 bp in length
 * 99576 99675: gap of 100 bp
 * 99676 104567: contig of 4892 bp in length
 * 104568 104667: gap of 100 bp
 * 104668 109346: contig of 4679 bp in length
 * 109347 109446: gap of 100 bp
 * 109447 116807: contig of 7361 bp in length.

FEATURES

source

1. 116807
 /organism="Danio rerio"
 /mol_type="genomic DNA"
 /db_xref="taxon:7955"
 /clone="DKEY-46N5"
 /clone_lib="DanioKey"
 1. 2116
 /note="assembly fragment:00043
 fragment_chain:1"
 2217. 8578
 /note="assembly fragment:00104
 fragment_chain:1"
 8679. 14990
 /note="assembly fragment:00322
 fragment_chain:1"
 15091. 28940
 /note="assembly fragment:00769
 fragment_chain:1"
 27041. 34475
 /note="assembly fragment:00537
 fragment_chain:1"
 34576. 41111
 /note="assembly fragment:00392
 fragment_chain:1"
 41212. 52851
 /note="assembly fragment:00633
 fragment_chain:2"
 52952. 58237
 /note="assembly fragment:00201
 fragment_chain:2"
 58338. 73639
 /note="assembly fragment:00925
 fragment_chain:2"
 73740. 96363
 /note="assembly fragment:01125
 fragment_chain:2"
 96464. 99575
 /note="assembly fragment:00066.0"
 99676. 104567
 /note="assembly fragment:00151"
 104668. 109346
 /note="assembly fragment:00258"
 109447. 116807
 /note="assembly fragment:00464"

ORIGIN

Query Match 5.7%; Score 86; DB 2; Length 116807;
 Best local Similarity 44.7%; Pred. No. 3.1e-05;
 Matches 335; Conservative 0; Mismatches 415; Indels 0; Gaps 0;
 QY 280 ACTACTAATGAGGAGCAATCTATGTAACCTAAAGAGTACTCGGTCCAGCAATGAGATT 339
 Db |||||
 26253 AATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 26194
 QY 340 GCGCATGGGCTAAATCTATACGCTAATACACTACACAGTTTCCCACTGTTACTGAGTCA 399
 Db |||||
 26193 GATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 26134
 QY 400 AATAATGATGGTACTGAGGTATTAATGTTCCCAATATGATATTAATGTTTCTAGC 459
 Db |||||
 26133 AATAAATAATGATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAAAGAAAT 26074
 QY 460 ACTGTTAATAATGAGCTGTAATATGTTTACATCTCTGTAACCTCCAAATGCTACTATTTCAT 519

Db 26073 AATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATGAT 26014
 QY 520 GAAAAGAAATACTGATCGACATGGGAGATGGTGGGAAAAAAGTGTAGATCAAAAAACG 579
 Db 26013 AATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 25954
 QY 580 TACTCGGTTGGTATACAGTCAAAATATCTATTACTTATTAAGAAATGCGAGTCAATTATCAT 639
 Db 25953 AATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 25894
 QY 640 GTTACAGAAAAAGTGTATCAATATGTTTATAAGGATACATGCGCATCTGCTTCTGTAGTT 699
 Db 25893 AATGATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 25834
 QY 700 GATTTGAACGAAGGCTTTATGAGTAACTATTACTTACTGATGATGATGATGATGATGATGAT 759
 Db 25833 AATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATGAT 25774
 QY 760 CTAACCAAGTTTCGGAAGCAACTGGGAAGTATAACCTGTTAGAGGAAAAATAATAAT 819
 Db 25773 AATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 25714
 QY 820 TTACGATTTACTATTCGGTGGGAGCTACCAATACCTCCAAACCGAAATACTCAAAATGGA 879
 Db 25713 GATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATGAT 25654
 QY 880 GCTAATGATGACTTTTTTTTAAAGGAATAATAATAATAATAATAATAATAATAATAATAAT 939
 Db 25653 AATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 25594
 QY 940 TTAAGAGTGGAGCTAAACCGAGTTTCAGCTGATTTTACCAGAAAAATAACAACATTTGGAC 999
 Db 25593 GATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATGAT 25534
 QY 1000 ATCAACCCCAATCTAGCAATGATGACCCA 1029
 Db 25533 AATAATGATAATAATAATAACAAAAACAACA 25504

AL935272 146275 bp DNA linear HTG 30-OCT-2003
 Danio rerio clone CH211-266K22, WORKING DRAFT SEQUENCE, 2 unordered pieces.
 AL935272
 AL935272.16 GI:37936458
 HTG; HTGS PHASE1; HTGS ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
 Danio rerio (zebrafish)
 Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
 1 (bases 1 to 146275)
 Andrew.R.
 Direct Submission
 Submitted (29-OCT-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Oct 23, 2003 this sequence version replaced gi:35763272.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: zfish-help@sanger.ac.uk
 ----- Project Information
 Center project name: zC266K22
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Chemistry: Dye-terminator; 100% of reads
 Consensus quality: 145872 bases at least Q40
 Consensus quality: 145995 bases at least Q30
 Consensus quality: 146056 bases at least Q20

RESULT 8
 AL935272
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT


```
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEY-245C12"
/clone_lib="DanicKey"
1..55329
/note="assembly_fragment:03483
fragment_chain:1"
55430..190714
/note="assembly_fragment:01158
fragment_chain:1"
190815..196436
/note="assembly_fragment:02910"
196597..198743
/note="assembly_fragment:03681"

misc_feature
1..55329
/note="assembly_fragment:03483
fragment_chain:1"
55430..190714
/note="assembly_fragment:01158
fragment_chain:1"
190815..196436
/note="assembly_fragment:02910"
196597..198743
/note="assembly_fragment:03681"

ORIGIN
Query Match 5.6%; Score 84; DB 2; Length 198743;
Best Local Similarity 45.5%; Pred. No. 5.5e-05;
Matches 300; Conservative 0; Mismatches 360; Indels 0; Gaps 0;

Qy 243 TAAAGCTTCAACTGATTTTAAATCTCTTTTACGACAACTACTAATGAGGGAGAACATA 302
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 86200 TACTACTACTACTTATAATATATATATAATATAATATAATATAATATAATATAATA 86141
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 303 TGTAACTAAAGAGATCTGGCTGAGCAATGAGATTGCGACATGGGCTAAATCTATATC 362
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 86140 TGATAATAATAATAATAATATGATAAAATGATAATAATAATAATGATAATAATAA 86081
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 363 AGCTAATACTACACCGAGTTCCACTGTTACTGAGTCAAATAATGATGCTAGGTTAT 422
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 86080 TGATAATAATAATAATAATGATAATAATAATAATAATAATAATAATAATGATAATA 86021
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 423 TAATGTTTCCCAATGATGATTTATTTCTTTAGCACCTGTTAAATGAGCTGTAAT 482
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 86020 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATGATAATA 85961
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 483 TATGGTTACCTCTGTAATCCAAATGCTATTTTCATGAGAAAGAAATGCTGATGCGCATG 542
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 85960 TAATGATAATAATAATAATGATAATAATAATAATAATAATAATAATAATAATAATA 85901
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 543 GGGAGATGGTGGTGGAAAACTGATGATCAAAAACGCTACTCGGTGGTGTATCAAGTCA 602
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 85900 TAATGATAATAATAATAATGATAATAATAATAATAATAATAATAATAATAATAATA 85841
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 603 ATATATATTTACTTATTAAGATGCGATCAATTAATCATGGTACAGAAAAAGTGTATCAATA 662
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 85840 TGATAATAATAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATA 85781
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 663 TGTATAAGGATCTATGCTCTCTCTGTTGATTTGACGAAAGGCTTTATGA 722
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 85780 TGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 85721
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 723 AGTAATCTATTACTGATGGATCAGGGAATATTACAACTCTAACTCAAGTTTCGAAAAAGC 782
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 85720 TAATAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 85661
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 783 AACTGGGAAGTATTAACCTGTTAGGAGAAATAATAATTTACGATTTACTATTCGTTGGGC 842
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 85660 TAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 85601
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 843 AGCTACCAATCTCCACCGGAATACTCAAAATGGAGCTAAATCATGATGATGATGATGAT 902
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 85600 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 85541
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10
AC007926/c 155204 bp DNA linear HTG 17-JUL-2001
LOCUS AC007926
DEFINITION Trypanosoma brucei chromosome II clone RPCI93-3H15, *** SEQUENCING
IN PROGRESS ***.
ACCESSION AC007926
VERSION AC007926.8 GI:14787210
KEYWORDS HTG; HTGS_PHASE2.
```

```
Trypanosoma brucei
Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
1 (bases 1 to 155204)
El-Sayed,N.M., Ghedin,E., Song,J., Larkin,C., Wanless,D., Jones,K.,
Peterson,J., Hou,L., Zhao,H., Mason,T., Militescher,J., Pai,G., Van
Aken,S., Utterback,T., Khalak,H.G., Gerard,C., Leech,V., Ullu,E.,
Melville,S., White,O., Adams,M.D., Donelson,J.E. and Fraser,C.M.
Trypanosoma brucei GUTat10.1 RPCI93-3H15 BAC genomic sequence
Unpublished
2 (bases 1 to 155204)
El-Sayed,N.M., Khalak,H. and Adams,M.D.
Direct Submission
Submitted (28-JUN-1999) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 155204: contig of 155204 bp in length.
FEATURES
Location/Qualifiers
1..155204
/organism="Trypanosoma brucei"
/mol_type="genomic DNA"
/isolate="GUTat10.1"
/db_xref="taxon:5691"
/chromosome="II"
/clone="RPCI93-3H15"
ORIGIN
Query Match 5.5%; Score 83.4; DB 2; Length 155204;
Best Local Similarity 44.6%; Pred. No. 7.5e-05;
Matches 327; Conservative 0; Mismatches 406; Indels 0; Gaps 0;

Qy 170 ATAAATGCAATGATCTGATTCGATTAAGATAAGATGGAGCTTCTTATTTAAATTCCTCAAGGTA 229
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 81317 ATAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 81258
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 230 AAGAAGCTGAGTATAAGCTTCAACTGATTTTAAATTCCTTTTACGACAACTACTAATG 289
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 81257 ATAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 81198
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 290 GAGGGAGAACATATGTAACATAAAAAAGATACGCGTCAGCAAAATGAGATTGCGACATGGG 349
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 81197 ATAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 81138
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 350 CTAAATCTATATCAGCTAATCTACACCAAGTTCCACTGTTACTGAGTCAATAATGATG 409
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 81137 ATAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 81078
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 410 GTACTCAGGTATTAATGTTTCCCAATATGGATATTATTATGTTTCTAGCACTGTTAAATA 469
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 81077 ATAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 81018
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 470 ATGGAGCTGTAATTAATGTTTACATCTGTAACTCCAAATGCTACTATTCATGAAAGAAATA 529
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 81017 ATAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 80958
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 530 CTGATCGCATGGGGAGATGGTGGGAAACCTGATAGTCAAAAACGCTACTCGGTG 599
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 80957 ATAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 80898
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 590 GTGATACAGTCAAAATATCTATTACTTTAAGAATGCGAGTCAATTTATCATGTCAGAGAA 649
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 80897 ATAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 80838
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

QY 650 AAGTGATCAATGTTTAAAGGATGACTATGCCTCTCTGTTAGTTGTAACG 709
Db 80837 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA
QY 710 AAGGCTCTTATGAAGTAACCTTACTGATGAGTACAGGAATATTACAACTCAACTCAAG 769
Db 80777 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA
QY 770 GTTCGGAAGAACCTGGGAAGTATAACCTGTTAGAGGAATAATAATAATTCACGATTA 829
Db 80717 ATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA
QY 830 CTATTCGCTGGGAGCTACCAATCTCCAAACCGGAAATCTCAAAATGGAGCTAATGATG 889
Db 80657 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA
QY 890 ACTTTTTTTTATAA 902
Db 80597 ATGATAATAATAA 80585

RESULT 11

AL929250/c

LOCUS

DEFINITION

Zebrafish DNA sequence from clone CH211-137A2, complete sequence.

ACCESSION

AL929250

VERSION

AL929250.8 GI:34787249

KEYWORDS

HTG.

SOURCE

Danio rerio (zebrafish)

ORGANISM

Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Ostariophysi;

Cypriniformes; Cyprinidae; Danio.

1 (bases 1 to 94534)

Almeida.J.

Direct Submission

Submitted (16-SEP-2003) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

zf1sh-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Sep 16, 2003 this sequence version replaced gi:27848021.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Web site: http://www.sanger.ac.uk

Contact: zf1sh-help@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the

assembly was confirmed by restriction digest, except on the rare

occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information

on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep CH211-137A2 is

from a CHORI-211 BAC library

VECTOR: pTRABAC2.1

Clone-derived Zebrafish pUC subclones occasionally display

inconsistency over the length of mononucleotide A/T runs and

conserved TA repeats. Where this is found the longest good quality

representation will be submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat

discovery system (Zhiyong Bao and Sean Eddy, submitted), and those

FEATURES

source

1. .94534

/organism="Danio rerio"

/mol_type="genomic DNA"

/db_xref="taxon:7955"

/clone="CH211-137A2"

/clone_lib="CHORI-211"

ORIGIN

Query Match

5.5%; Score 82.4; DB 5; Length 94534;

Best Local Similarity

44.5%; Pred.No.0.00013;

Matches 326; Conservative

0; Mismatches 406; Indels

0; Gaps

0;

QY 103 ATTACAGTTTCAAGATACTCAAAAAGGCGCAACCTATAAAAGCATATAAAGTTTTTGATGCA 162

Db 30450 AATATAGTTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA

QY 163 GAAATAGATAATGCAAAATGATCTGATTGCAATAAAGATGGAGCTTCTTTAATTTTCCT 222

Db 30390 AATAATATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT

QY 223 CAAGGTAAAGAACTGAGTATAAGCTTCAACTGATTTTAATTTCTCTTTTACGACAACT 282

Db 30330 AATAATAATAATGATGATAATAAAGATAATAATAATAATAATAATAATAATAATAATA

QY 283 ACTAATGGAGGAGAAACATATGTAACTAAAAAGAGATACTCGTCAGCAAAATGAGATTGCG 342

Db 30270 AATAATGTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA

QY 343 ACATGGGCTAAATCTATATCAGCTAATACTACACCACTTTCCACTGTTTACTGAGTCAAAAT 402

Db 30210 AATAATAGTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA

QY 403 AATGATGGTACTGAGGTATTATTGTTCCCAATATGATATTATTATTATTATTATTATTATT 462

Db 30150 AATAATGGTGATGATGATAATAATAATGATAATAATAATAATAATAATAATAATAATAA

QY 463 GTTAATAATGGAGCTGTAAATTTATGTTTACATCTGTAATCTCCAAATGCTACTATTTCATGAA 522

Db 30090 AGTAATAATAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA

QY 523 AAGAATACTGATCGGACATCGGGAGATGGTGGGAAAACTGTAGATCAAAAAACCTAC 582

Db 30030 AATAATAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA

QY 583 TCGGTTGGTGATACAGTCAATATATCTATTACTTTAAGAAATGCAGTCAATTTATCATGGT 642

Db 29970 AATAATAATGGTGATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAA

QY 643 ACAGAAAAAGTGTATCAATATGTTTATAAAGGATGATGCTCTGCTCTGTAGTTGAT 702

Db 29910 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA

QY 703 TTGAACCAAGGGCTTTATGAGTAACCTATTACTTGATGGATCAGGGAATATTACAACTCTA 762

Db 29850 AGTAATAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA

QY 763 ACTCAAGGTTCCGAAAAAGCAACTGGGAGTATAAAGCTGTTAGAGGAAATAATAATTTTC 822

Db 29790 AATAATAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA

QY 823 ACCATTACTATT 834

Db 29730 AATAATTTAATA 29719

RESULT 12

PFMAL13.07

WPCOMMENT

Sequence split into 28 fragments LOCUS PFMAL13 Accession AL844509

Db 43595 ATAATAATCATATAATAATAATTATTTATTAATAATAACAAAGAAAAAAGCGAAAAAG 43654

QY 737 ATGGATCAGGAATATTACAACCTCAACTCAAGGTTCCGAAAAGCAACTCGGAAGTATA 796

Db 43655 AAGAAGTATCATTTATATATCTGAACCCCAATGCTATAGAACTAATATATGGTAATT 43714

QY 797 ACCTGTTAGAGAAAATAATAATTTC 823

Db 43715 ATATTCCCAGGATGAAGATAATTACA 43741

RESULT 13
BX537105
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BX537105 225581 bp DNA linear HTG 06-AUG-2003
Danio rerio clone DKEY-3L24, 18 unordered pieces.
BX537105
BX537105.2 GI:31559379
HTG; HTGS PHASE1; HTGS CANCELLED.
Danio rerio (zebrafish)
Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 225581)
Burton,J.
Direct Submission
Submitted (05-AUG-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jun 9, 2003 this sequence version replaced gi:31441851.

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: zk3L24
----- Summary Statistics
Assembly program: XCAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 217427 bases at least Q40
Consensus quality: 219130 bases at least Q30
Consensus quality: 220818 bases at least Q20
Insert size: 223881; sum-of-contigs
Insert size: 219000; 7.3% error; agarose-fp
Quality coverage: 6.82x in Q20 bases; agarose-fp
coverage: 7.43x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 20602: contig of 20602 bp in length
* 20603 20702: gap of 100 bp
* 20703 57831: contig of 37129 bp in length
* 57832 57931: gap of 100 bp
* 57932 65164: contig of 7233 bp in length
* 65165 65264: gap of 100 bp
* 65265 71392: contig of 6128 bp in length
* 71393 71492: gap of 100 bp
* 71493 93677: contig of 22185 bp in length
* 93678 93777: gap of 100 bp
* 93778 101598: contig of 7821 bp in length
* 101599 101698: gap of 100 bp
* 101699 128689: contig of 26991 bp in length
* 128690 128789: gap of 100 bp
* 128790 138632: contig of 9843 bp in length
* 138633 138732: gap of 100 bp
* 138733 144193: contig of 5461 bp in length

QY	965	CAGCTGATTACCAAGAAATACAAACATTCGACCATCAACCCCAATACTAGCAATGATG	1024
Db	166410	ATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAG	166351
QY	1025	ACCCAGGTCAAAAAGTAAACAGTGAGGGATGGTCAAATTTACTATATAAAAAAATTGATGGT	1083
Db	166350	ATACAGATACAGATACAGGATTTGGAGCTGGAGTCATTTAAAGGAGTCAAGAGTGCAGGT	166292

Search completed: September 13, 2004, 08:14:20
Job time : 5991.67 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2004, 10:28:56 ; Search time 45 Seconds
(without alignments)
3493.409 Million cell updates/sec

Title: US-10-009-254-2

Perfect score: 2578

Sequence: 1 MKKMIQSLVSLVSLAFGKAV.....GALLVIGIVLVARRLRS 502

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1288442 seqs, 313154207 residues

Total number of hits satisfying chosen parameters: 1288442

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database :

Published Applications AA.*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCR_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2578	100.0	502	12	US-10-333-002-16
2	334	13.0	627	12	US-10-206-576-218
3	334	13.0	627	16	US-10-661-809-13
4	318	12.3	625	16	US-10-661-809-19
5	274.5	10.6	560	12	US-10-206-576-220
6	265.5	10.3	665	10	US-09-769-787-127
7	232	9.0	793	12	US-10-282-122A-60689
8	222.5	8.6	688	9	US-09-864-761-36047
9	220	8.5	560	12	US-10-333-002-19
10	219	8.5	915	14	US-10-193-764-35
11	219	8.5	1222	14	US-10-193-764-37
12	219	8.5	1228	14	US-10-193-764-34
13	218	8.5	1465	15	US-10-369-493-13955
14	212	8.2	645	12	US-10-282-122A-73634
15	212	8.2	2551	12	US-10-282-122A-74083
					Sequence 16, Appl
					Sequence 219, Appl
					Sequence 13, Appl
					Sequence 19, Appl
					Sequence 32, Appl
					Sequence 127, Appl
					Sequence 60689
					Sequence 36047, A
					Sequence 19, Appl
					Sequence 35, Appl
					Sequence 37, Appl
					Sequence 34, Appl
					Sequence 13955, A
					Sequence 73634, A
					Sequence 74083, A

16	208.5	8.1	476	16	US-10-661-809-11	Sequence 11, Appl
17	200.5	7.8	886	10	US-09-769-787-126	Sequence 126, Appl
18	199	7.7	473	16	US-10-661-809-17	Sequence 17, Appl
19	198.5	7.7	724	12	US-10-282-122A-45795	Sequence 45795, A
20	198	7.7	596	10	US-09-946-374-243	Sequence 243, Appl
21	198	7.7	596	12	US-10-206-915-310	Sequence 310, Appl
22	198	7.7	596	12	US-10-199-670-310	Sequence 310, Appl
23	198	7.7	596	12	US-10-201-858-310	Sequence 310, Appl
24	198	7.7	596	12	US-10-205-890-310	Sequence 310, Appl
25	198	7.7	596	12	US-10-208-024-310	Sequence 310, Appl
26	198	7.7	596	12	US-10-201-853-310	Sequence 310, Appl
27	198	7.7	596	12	US-10-063-745-100	Sequence 100, Appl
28	198	7.7	596	12	US-10-063-512-100	Sequence 100, Appl
29	198	7.7	596	12	US-10-063-513-100	Sequence 100, Appl
30	198	7.7	596	12	US-10-063-515-100	Sequence 100, Appl
31	198	7.7	596	12	US-10-063-549-100	Sequence 100, Appl
32	198	7.7	596	12	US-10-063-569-100	Sequence 100, Appl
33	198	7.7	596	12	US-10-063-551-100	Sequence 100, Appl
34	198	7.7	596	12	US-10-174-581-310	Sequence 310, Appl
35	198	7.7	596	12	US-10-176-483-310	Sequence 310, Appl
36	198	7.7	596	12	US-10-176-749-310	Sequence 310, Appl
37	198	7.7	596	12	US-10-176-914-310	Sequence 310, Appl
38	198	7.7	596	12	US-10-176-915-310	Sequence 310, Appl
39	198	7.7	596	12	US-10-006-485A-243	Sequence 243, Appl
40	198	7.7	596	12	US-10-013-907A-243	Sequence 243, Appl
41	198	7.7	596	12	US-10-015-499A-243	Sequence 243, Appl
42	198	7.7	596	12	US-10-063-555-100	Sequence 100, Appl
43	198	7.7	596	12	US-10-063-563-100	Sequence 100, Appl
44	198	7.7	596	12	US-10-063-594-100	Sequence 100, Appl
45	198	7.7	596	12	US-10-063-553-100	Sequence 100, Appl
46	198	7.7	596	12	US-10-063-554-100	Sequence 100, Appl
47	198	7.7	596	12	US-10-176-484-310	Sequence 310, Appl
48	198	7.7	596	12	US-10-180-550-310	Sequence 310, Appl
49	198	7.7	596	12	US-10-183-014-310	Sequence 310, Appl
50	198	7.7	596	12	US-10-187-738-310	Sequence 310, Appl
51	198	7.7	596	12	US-10-187-740-310	Sequence 310, Appl
52	198	7.7	596	12	US-10-187-683-310	Sequence 310, Appl
53	198	7.7	596	12	US-10-194-363-310	Sequence 310, Appl
54	198	7.7	596	12	US-10-194-460-310	Sequence 310, Appl
55	198	7.7	596	12	US-10-194-463-310	Sequence 310, Appl
56	198	7.7	596	12	US-10-194-484-310	Sequence 310, Appl
57	198	7.7	596	12	US-10-195-884-310	Sequence 310, Appl
58	198	7.7	596	12	US-10-195-896-310	Sequence 310, Appl
59	198	7.7	596	12	US-10-196-744-310	Sequence 310, Appl
60	198	7.7	596	12	US-10-196-755-310	Sequence 310, Appl
61	198	7.7	596	12	US-10-196-757-310	Sequence 310, Appl
62	198	7.7	596	12	US-10-197-704-310	Sequence 310, Appl
63	198	7.7	596	12	US-10-197-710-310	Sequence 310, Appl
64	198	7.7	596	12	US-10-198-758-310	Sequence 310, Appl
65	198	7.7	596	12	US-10-198-766-310	Sequence 310, Appl
66	198	7.7	596	12	US-10-199-304-310	Sequence 310, Appl
67	198	7.7	596	12	US-10-199-309-310	Sequence 310, Appl
68	198	7.7	596	12	US-10-199-313-310	Sequence 310, Appl
69	198	7.7	596	12	US-10-199-456-310	Sequence 310, Appl
70	198	7.7	596	12	US-10-201-329-310	Sequence 310, Appl
71	198	7.7	596	12	US-10-202-412-310	Sequence 310, Appl
72	198	7.7	596	12	US-10-206-919-310	Sequence 310, Appl
73	198	7.7	596	12	US-10-206-922-310	Sequence 310, Appl
74	198	7.7	596	12	US-10-206-924-310	Sequence 310, Appl
75	198	7.7	596	12	US-10-206-928-310	Sequence 310, Appl
76	198	7.7	596	12	US-10-207-314-310	Sequence 310, Appl
77	198	7.7	596	12	US-10-207-921-310	Sequence 310, Appl
78	198	7.7	596	12	US-10-207-922-310	Sequence 310, Appl
79	198	7.7	596	12	US-10-208-027-310	Sequence 310, Appl
80	198	7.7	596	12	US-10-013-910A-243	Sequence 243, Appl
81	198	7.7	596	12	US-10-174-570-310	Sequence 310, Appl
82	198	7.7	596	12	US-10-183-005-310	Sequence 310, Appl
83	198	7.7	596	12	US-10-236-254A-243	Sequence 243, Appl
84	198	7.7	596	12	US-10-015-395A-243	Sequence 243, Appl
85	198	7.7	596	13	US-10-006-867-100	Sequence 100, Appl
86	198	7.7	596	13	US-10-052-586-310	Sequence 310, Appl
87	198	7.7	596	13	US-10-063-547-100	Sequence 100, Appl
88	198	7.7	596	14	US-10-174-590-310	Sequence 310, Appl

89 198 7.7 596 14 US-10-176-758-310 Sequence 310, App
90 198 7.7 596 14 US-10-175-737-310 Sequence 310, App

ALIGNMENTS

RESULT 1
US-10-333-002-16
; Sequence 16, Application US/10333002
; Publication No. US20040071729A1
; GENERAL INFORMATION:
; APPLICANT: Adderson, Elisabeth
; APPLICANT: Bohnsack, John
; TITLE OF INVENTION: Group B Streptococcus Polypeptides Nucleic Acids and Therapeutic
; TITLE OF INVENTION: Compositions and Vaccines Thereof
; FILE REFERENCE: 2511-1-001 (SJ-0039)
; CURRENT APPLICATION NUMBER: US/10/333,002
; CURRENT FILING DATE: 2003-07-08
; PRIOR APPLICATION NUMBER: PCT/US01/24795
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 09/634,341
; PRIOR FILING DATE: 2000-08-08
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Streptococcus agalactiae
US-10-333-002-16

Query Match 100.0%; Score 2578; DB 12; Length 502;
Best Local Similarity 100.0%; Pred. No. 1.3e-198;
Matches 502; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKKKMIQSLVLASLAFGMVSPVTPPIAFAAETGTTTVDQTKGATYKAYKVFDAEIDNAN 60
DB 1 MKKKMIQSLVLASLAFGMVSPVTPPIAFAAETGTTTVDQTKGATYKAYKVFDAEIDNAN 60
QY 61 VSDSNKDGASLYLPQKGEABYKASTDFNSLPTTTTNGGRTYVTKDTASANEATWAKSI 120
DB 61 VSDSNKDGASLYLPQKGEABYKASTDFNSLPTTTTNGGRTYVTKDTASANEATWAKSI 120
QY 121 SANTTPVSTVSTESNNDGTEVINVSQYGYVSVSTVANGAVIMVTSVTPNATHEKNTDAT 180
DB 121 SANTTPVSTVSTESNNDGTEVINVSQYGYVSVSTVANGAVIMVTSVTPNATHEKNTDAT 180
QY 181 WGDGKGKTVDPQKTSYSGDTPVKYITTYKNAVNYHGTEKTYQYVTKDTMPASVVDLNEGSY 240
DB 181 WGDGKGKTVDPQKTSYSGDTPVKYITTYKNAVNYHGTEKTYQYVTKDTMPASVVDLNEGSY 240
QY 241 EVTITDGSNGITTLTGSEKATGKYNLENNFTTIPWAAATNTPTGNTONGANDDFY 300
DB 241 EVTITDGSNGITTLTGSEKATGKYNLENNFTTIPWAAATNTPTGNTONGANDDFY 300
QY 301 KGINTITVYTGVLKSGKPGSADLPENTNIATINPNTSNDPQKQVTVRVDGQITIKKID 360
DB 301 KGINTITVYTGVLKSGKPGSADLPENTNIATINPNTSNDPQKQVTVRVDGQITIKKID 360
QY 361 GSTKASLQGAIFVLKNAQGLFNFNDTNVNWEGTEANATEYTTGADGIIITLKGEGTY 420
DB 361 GSTKASLQGAIFVLKNAQGLFNFNDTNVNWEGTEANATEYTTGADGIIITLKGEGTY 420
QY 421 LVEKAPLGNLNDNSQKVLGQADTNTSDNLLNVPNTVENNKGTELSTGGIGTIFY 480
DB 421 LVEKAPLGNLNDNSQKVLGQADTNTSDNLLNVPNTVENNKGTELSTGGIGTIFY 480
QY 481 IIGAILVIGAGIVLVARRRLLS 502
DB 481 IIGAILVIGAGIVLVARRRLLS 502

RESULT 2

US-10-206-576-218
; Sequence 218, Application US/10206576
; Publication No. US20030017495A1
; GENERAL INFORMATION:
; APPLICANT: Choi et al.
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 497
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-R
; COMPUTER: Dell Latitude
; OPERATING SYSTEM: Windows 98
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/206,576
; FILING DATE: 29-Jul-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/071,035
; FILING DATE: 1998-05-04
; APPLICATION NUMBER: US 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: US 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: US 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Hyman, Mark J.
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB369P1D1
; INFORMATION FOR SEQ ID NO: 218:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 627 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 218:
US-10-206-576-218

Query Match 13.0%; Score 334; DB 12; Length 627;
Best Local Similarity 26.3%; Pred. No. 5e-18;
Matches 150; Conservative .79; Mismatches 223; Indels 118; Gaps 26;
QY 3 KKMIOQLVLASLAFGMVSPVTPPIAFAA--ETGTTTVDQTKGATYKAYKVFDAEIDNAN 60
DB 100 KQAVQSL-----TPGKPVAGITDANGVTVQLPKKQNGKDAVYTIKEPKGV 148
QY 61 VSDSNKDGAA--SYLIPQKGEABYKASTDFNSLP-----TTTTNGGRTYVTKDTASANE- 112
DB 149 VAATNVAWAFVYEMIKOTDGSYKGTBELAVHYHPKNVAVANDGSLHVKYGTAE--NEG 207
QY 113 -----IATWA-----KSIANTTPV--STVTESNND 136
DB 208 LNGAEFVISKSGSPGVKVIQGVKGLTWTDTKEQAKRFITGKSYEIGENDPTEAEN- 266
QY 137 GTEVINVS--QYGYTVVS--STVNNGAVIMVTSVTP-----NATHEK--NTDATWGD 184
DB 267 GTGELTVKNLEVGSYILEBVKAPNNAELIENQTKTPTIEANNQTPVEKTVKNDTSKVDK 326
QY 185 GSKTVDPQKTSYSGDTPVKYITTYKNAVNYHGTE-----KVYQYVTKDTMPASVVD--LNEGS 239
DB 327 TTPSLDGDQDAIGEKIKYQISVNIPLGIADKGDANKYKFNLDKHDALTFDNYTSGE 386
QY 240 YEVTITDGSNGITTLTGSEKATGKYNLENNFTTITI--PWAATNTPTGNTONGANDD 297
DB 387 YAYALYDGDTVI-----APENYQVTEQANGFTVAVNPAYIPTLTFGG----- 428